2010-09

Assessing neurophysiologic markers for training and simulation to develop expertise in complex cognitive tasks

Sullivan, Joseph A.

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

ASSESSING NEUROPHYSIOLOGIC MARKERS FOR TRAINING AND SIMULATION TO DEVELOP EXPERTISE IN COMPLEX COGNITIVE SKILLS

by

Joseph A. Sullivan

September 2010

Dissertation Supervisor: Rudolph P. Darken

This dissertation was done at the MOVES Institute
Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE | *Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 2010	Dissertation

4. TITLE AND SUBTITLE Assessing Neurophysiologic Markers for Training and Simulation to Develop Expertise in Complex Cognitive Tasks

5. FUNDING NUMBERS

6. AUTHOR(S) Joseph A. Sullivan

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number NPS.2010.0043-IR-EP7-A.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

This work explores the theoretic basis and provides empirical support for using neurophysiologic markers to provide information on a trainee's cognition to guide instruction. This serves as the basis for improving the design of simulation responsive to individual traits for training continuous complex cognitive tasks. Individualized instruction has been empirically proven to be vastly superior to other forms of instruction. However, current methods to design simulation that is responsive to the user have relied primarily on raw performance metrics. These metrics are often misleading and provide very little diagnostic value. For complex tasks, understanding cognitive processes is critical. Neurophysiologic markers can potentially inform instructional systems on trainees' cognition but have yet to be validated. This research developed a sample process to identify neurophysiologic markers for informing individualized instruction. Applying the process to helicopter overland navigation, a theoretic model of eye scan behavior was developed. The process and theoretic model were validated by analyzing novices and expert navigators. Predicted eye scan metrics reliably distinguished between expert and novice behavior, providing insight not available using raw performance metrics. Also, a visualization tool was developed to explore expert scan strategies. In addition to confirming expected strategies and novice expert differences, we discovered novel, unexpected strategies of expert navigators.

14. SUBJECT TERMS Training, Simulation, Human Factors, Aviation, Helicopter, Navigation, Neurophysiology, Expertise

15. NUMBER OF PAGES
181

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500 | Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited.

ASSESSING NEUROPHYSIOLOGIC MARKERS FOR TRAINING AND SIMULATION TO DEVELOP EXPERTISE IN COMPLEX COGNITIVE TASKS

Joseph A. Sullivan
Commander, United States Navy
B.S., Catholic University of America, 1986
M.S., Naval Postgraduate School, 1998

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN
MODELING, VIRTUAL ENVIRONMENTS AND SIMULATION (MOVES)

from the

NAVAL POSTGRADUATE SCHOOL
September 2010

Author: ____
Joseph A. Sullivan

Approved by:

Rudolph Darken Professor of Computer Science Dissertation Supervisor	Anthony Ciavarelli Research Professor of MOVES
LCDR Joseph Cohn, PhD Instructor of MOVES	Theodore Lewis Professor of Computer Science
Michael McCauley Research Professor of Operations Research	Gurminder Singh Professor of Computer Science

Approved by: ____
Mathias Kölsch, Chair, MOVES Academic Committee
Peter J. Denning, Chair, Computer Science Department

Approved by: ____
Douglas Moses, Vice Provost for Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This work explores the theoretic basis and provides empirical support for using neurophysiologic markers to provide information on a trainee's cognition. Improved insight into cognition serves as the basis for improving the design of simulation responsive to individual traits for training continuous complex cognitive tasks. Individualized instruction has been empirically proven to be vastly superior to other forms of instruction. However, current methods to design simulation that is responsive to the user have relied primarily on raw performance metrics. These metrics are often misleading and provide very little diagnostic value. For complex tasks, understanding cognitive processes is critical. Neurophysiologic markers can potentially inform instructional systems on trainees' cognition but have yet to be validated. This research developed a sample process to identify neurophysiologic markers for informing individualized instruction. Applying the process to helicopter overland navigation, a theoretic model of eye scan behavior was developed. The process and theoretic model were validated by analyzing novices and expert navigators. Predicted eye scan metrics reliably distinguished between expert and novice behavior, providing insight not available using raw performance metrics. Also, a visualization tool was developed to explore expert scan strategies. In addition to confirming expected strategies and novice expert differences, we discovered novel, unexpected strategies of expert navigators.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGEMENTS

This dissertation would not have been possible without tremendous help from an amazing array of leaders, mentors, sponsors, colleagues, friends and family. My sincerest thanks to the Naval Postgraduate School leadership for its dedication to providing an educational environment for the military leaders of the future. In particular: ‘Thank You’ to Dr. Karl van Bibber; your personal involvement and encouragement provided the ideal catalyst, perfectly timed.

Few sponsors have been as supportive and effective as Mr. John Moore of the Navy Modeling and Simulation Office and Dr. Roy Stripling of the Office of Naval Research. John’s help goes well beyond the role of an outstanding research sponsor. He has been a great advocate for MOVES and NPS. Thank you, John, for your patient and unwavering support. Dr. Stripling continually encouraged us to expand our vision and extend our capabilities. Thank you, Roy for providing such exciting paths.

I could not have found a more talented or supportive dissertation committee. Special thanks to my advisor, Dr. Rudy Darken. Your leadership as the 2nd Director of MOVES, role as a mentor, support of my dissertation research, and friendship have been truly extraordinary. Thank you for excelling across the board in each of these roles.

Tremendous thanks to the Team in 212B, particularly: Dr. Ji Hyun Yang, Dr. Quinn Kennedy and Michael Day. You made difficult and onerous tasks look simple. Your enthusiasm and collegiality were a highlight of this work. Thank you, Ji Hyun, for countless, invaluable lessons and dedication to the highest-quality research.

The students, staff and faculty provide an amazing environment for generating innovative and practical solutions. Countless fellow researchers provided timely and helpful suggestions. My thanks also to friend and mentor, Paul Chatelier. Thank you for the encouragement, sage advice, and for believing this could be done.

Finally, and most importantly, I would like to thank my wife and best friend, Katie, and our wonderful family. You provided more support than I could possibly have imagined. Thank you, Katie, Kayleigh, Nick, Claire, Patrick and Mary. This work and my love are dedicated to you.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The goal of this research is to improve our understanding of the theoretical basis for selecting, evaluating and applying measures and indicators of internal mental processes to assess trainee's acquisition of skill in continuous complex cognitive tasks. Complex cognitive tasks are pervasive in the defense and security communities. They involve all aspects of cognition, from attention, to working memory, to decision making, and have a temporal component that requires continuous self-assessment and planning. Typically, complex cognitive tasks are not only difficult to master, but they occur in operational environments that are not conducive to on-the-job training. Current training programs rely on repeated, direct and largely unfiltered exposure to these events in their full complexity at maximum operational pace. There is little to no opportunity to adapt pace or task complexity to the specific needs of an individual trainee.

Because of the prevalence, importance and difficulties associated with training complex cognitive tasks, the design, application and evaluation of simulation to provide affordable, safe and efficient practice are critical research topics. Historically, simulation design and effectiveness evaluation methods have focused on the degree of similarity with the real-world task (fidelity) and the degree of job performance transfer that can be evaluated and measured. The focus of the effort described here is on methods for improving how simulation can be applied to training by tapping unobservable performance attributes—eye-scan in this case. This work will examine how eye scan can be used to assess internal factors such as perception, attention management and mental workload that contribute to development of skill on continuous complex cognitive tasks. This research will also explore how that assessment can be related to the trainee's level of expertise. While beyond the scope of this dissertation, the eventual goal is to use this information to inform training interventions that accelerate learning and the development of expertise. The following anecdote is useful for framing the problem, providing an overview of the proposed solution and describing the expected contribution. Although this scenario itself is fictional, the individual elements of the case are drawn from the direct experiences of the author.

A. BACKGROUND STORY

At 90 knots and 120' above ground level inbound to the checkpoint labeled 'Indian Pass' on the 1:50,000 scale topographic map, it was clear to the instructor that the student was working hard and was keeping up. The instructor was confident that the student knew where he was and which way he was headed, but he was working very hard to maintain that awareness. The student's running narrative of the navigation features in view matched the 'navigation' portion of the mission brief reasonably well. In the narrow windows of time available for the instructor to observe the student's actions,[1] it looked like the trainee was doing the right things: shifting his scan frequently; looking for terrain features well ahead of and all around the aircraft. He seemed to be keeping up with his kneeboard card, which outlined headings, timings, and fuel for each leg of the navigation route. Likewise, he seemed to be keeping his map aligned with the direction of travel. His finger moved across the map at a steady pace, and he seemed to have a good approximation of his position. Reaching Indian Pass within a few seconds of planned timing, the student called for the instructor to follow the terrain in a left turn around peak 452 toward the next check point. Almost subconsciously he remembered to reset the clock to back up his terrain association navigation with dead reckoning. So far, the instructor pilot was comfortable that training objectives were being met.

On course and on timing for the next checkpoint, the instructor thought he sensed a change in the student as peak 452 went out of view. The only noticeable difference was that the flow of communication seemed to taper off. Of course, the instructor had seen this type of behavior before, when students were working especially hard. Sometimes silence was followed by a major 'ah-ha' moment; other times it meant the student was beyond saturated and it was time to break the task into more manageable elements. The instructor was certain that this was a critical time when asking even straightforward questions could throw some students into an unrecoverable maintenance cycle where all learning stops until the instructor reorients the student.

[1] On this type of training flight, the instructor has the aircraft controls and the student is navigating. Consequently, the instructor does not have complete freedom to observe the student at all times.

The aircraft passed a distinct rock formation within a few hundred meters of the planned route, still on timing and near the planned route. Did silence mean the student was working hard and about to make a critical connection, or did it mean the student was confused and couldn't find words to describe either his dilemma or the help he might need? Was the student intentionally offset from the course to get a better view to the next checkpoint, or was he unaware that he was slightly off course? The instructor was reminded of a student he flew with earlier in the week. During that flight, the instructor had assumed that pointing out the use of the orientation of an adjacent ridge as a channeling feature might be helpful. In the post-flight debrief, the student described how critical this advice had been in allowing him to see the terrain differently. He described how this advice had helped him make better sense of the contour map, allowing him to succeed at navigation and concentrate on other parts of the mission. Would the same advice work, or was today's student near a solution and simply needed a bit more time, or perhaps was simply overwhelmed?

Both the content and timing of instruction depended entirely on what was going on in the student's head. These processes included the navigation strategy the student had selected, how well they were executing the strategy, how sure they were in their solution, and of course, how much attention they could spare for conversation. If the IP knew the student was overwhelmed, he could gain altitude, point out some key features, and then try navigating from a higher altitude and slower airspeed to simplify the task. He had seen this strategy work in the past, but he also knew it didn't work well for every student and didn't prepare students for follow on tactics phase flights. The instructor's choices to maximize training were actually more straightforward if the student was outright overwhelmed than if he were struggling, but not necessarily productively struggling. Even a brief conversation to find out the strategy the student was using, no less how well the strategy was being applied, would place added burden and potentially distract the student enough to ruin what could otherwise be a key learning moment. In short, the instructor wished he could look inside the student's head so that he would know exactly what circumstance he was facing which would help him select the right intervention for this student at this moment.

Glancing at the student, the instructor decided to err on the side of caution. Assuming it would be better to ensure that the student knew where he was, the instructor asked the student what seemed like an innocuous question that would provide insight into how well the student was keeping up. "How's our timing to the next checkpoint?" Unfortunately, this distraction was exactly what the instructor had hoped to avoid. What was a distinctive rock formation to the instructor was not so distinctive to the student and had confused him. To the student, it looked remarkably similar to a feature in the draw adjacent to and nearly parallel with the planned route. For the last several minutes, he had been struggling to find terrain features that would help distinguish between the two draws. Had the IP known the nature of that uncertainty, he could have simply pointed out to the student that heading information could be used to distinguish the two draws. Instead, the student realized he had momentarily overlooked timing and now tried to make up for it. The student switched his attention to his kneeboard and timing information.

Realizing they were nearing timing for the next checkpoint, he let the instructor know he wasn't certain where they were. With the next checkpoint approaching rapidly, the instructor ran out of options and simply pointed out the bend in a creek bed that defined the next checkpoint. Once reoriented, they continued navigation. The instructor knew they hadn't gotten everything they could out of this part of the flight, but wasn't sure what he could have done differently. He made a mental note to do his best to recreate this moment in the post-flight debrief to recover what learning they could.

B. FRAMING THE PROBLEM

The anecdote in the previous section describes a common problem in training many complex skills. In order to provide the best feedback to the student in the right form at the right time, the instructor pilot needed some insight into processes internal to the trainee: how the student was sensing and perceiving the environment, focusing his attention and processing information. Was he lost? Was he momentarily disoriented? Was he knowingly altering the flight route for some reason? Was he on the verge of a significant breakthrough in understanding and ability? Depending on the answer, the instructor would provide different feedback in a different form. If the instructor knew

that a student was completely overwhelmed, he would likely stop and reorient the training in a simplified form or break it into smaller pieces. If the instructor knew a trainee was just beginning to master component skills such as correlating terrain features with the correct contour map representation, he would likely point out a wide range of increasingly difficult salient features. If the instructor knew that the student was on the right track but near mental workload limits, he would carefully time his input and suggestions so as not to interfere with emerging knowledge and skills.

It is inefficient and costly to rely on rote repetition or other “one-size-fits-all” approaches for developing expertise on difficult tasks that involve significant perceptual and cognitive elements. Simulation is an appealing training option for these difficult, perceptual and cognitive tasks. Ideally, simulation solutions should provide training opportunities tuned to each individual’s skill level and responsive in real time to trainee’s state. The quality, timeliness and appropriateness of individualized instruction in live or simulated training events depend on the accuracy of the instructor or instructional system’s assessment of trainee’s internal processes. However, simulation developers seldom implement solutions that automatically account for individual’s expertise level or respond in real time based on a trainee’s internal state. Live and simulated training systems rarely take advantage of indicators that are not human-observable.

As the dilemma presented in the introductory anecdote illustrates, it is not always obvious, even to experienced instructors, what set of cues and inputs should be considered as the basis for individualizing instruction. In live un-instrumented training, the instructor can never be certain of a student’s perception of the environment, attention management, mental processing and workload. How can an instructor or instructional system estimate largely internal processes that affect learning so dramatically? If these internal processes can be reliably assessed, what measures provide the best diagnostic value? For training highly cognitive tasks, it seems clear that tailored instruction requires the ability to make timely inferences about factors such as a trainee’s perception, attention, processing and workload. Currently, systems for sensing neurophysiologic data are rapidly advancing. Simultaneously, an emerging body of literature suggests that

this raw neurophysiologic data can be reliably turned into information on the trainee's internal state; and that this state information can provide diagnostic value to an instructional system.

This research project demonstrates that individualized training based solely on observable inputs (what the IP could see or extract verbally in the case of the anecdote) is limiting. This project also demonstrates that technologies that can monitor and measure unobservable inputs (eye tracking in this case, but electroencephalogram would be another such technology) can be used to more accurately model human performance and the development of expertise. This dissertation explores the theoretical basis for relying on unobservable cues, explores tradeoffs between the ability to sense and make use of indicators of internal processes, presents a process for selecting signals, and evaluates predicted and actual results.

C. MOTIVATION

Given current limits in science and technology, the majority of systems for training complex cognitive tasks are designed following a "one-size-fits all" approach. Even though the value of individualized instruction is widely recognized, it is not widely implemented. One of the key elements missing from the design process is reliable indicators of the trainee's perception of the environment and internal cognitive processes that could trigger individualized instruction. Without reliable insight into factors such as perception, attention and mental workload, the current training design process is focused on creating as near a literal recreation of the task as technology can support. Similarly, the evaluation process is focused on measuring real world task performance. The motivation of the present dissertation is to improve our ability to personalize instruction; extend our understanding of the methods for selecting, analyzing and applying neuromarkers for assessing internal processes associated with skill acquisition; provide insight into a trainee's performance on complex cognitive tasks; and improve our understanding of acquisition of spatial knowledge.

D. CONTRIBUTIONS

This research extends our current understanding of how neuromarkers are selected, and the utility of data they provide to assess critical processes internal to the trainee during the performance of a complex cognitive task. Previous work, reviewed in detail in Chapter II, established that post-hoc analysis of eye scan metrics is useful for distinguishing levels of experience at various psychomotor tasks. The research presented here extends previous investigation of neurophysiologic markers to explore a representative continuous complex cognitive task—helicopter overland navigation. We describe a process for creating a model of eye scan behavior that distinguishes between levels of expertise. We validate that model with experimental data collected in a simulation of the helicopter navigation task. Statistical analysis of eye tracking matched the model and predicts performance in the complex cognitive task of helicopter overland navigation. Visualization of individual's eye gaze patterns provides insight into trainee's proficiency not readily available in training or operational settings. The visualization capability was useful for classifying trainee's strategies and identifying training opportunities. The visualization capability also revealed un-expected strategies of more experienced participants. Thus, this work also extends our understanding of advanced navigation strategies and provides new insight into spatial knowledge acquisition during helicopter overland navigation using terrain association.

E. DISSERTATION OUTLINE

This dissertation is organized into the following chapters:

Background: This chapter provides requisite supportive material both for the overall context of where this research fits and for development of the experimental hypothesis. The background chapter covers the following topics that help develop the experimental hypothesis: human information processing, learning, models of expertise and spatial knowledge acquisition. In addition, the background covers the following topics that provide overall context for the overarching goals and potential impact of this research: instructional design methods, cognitive task analysis, scaffolding and part task training. Since application of neurophysiologic markers could also impact the simulation acceptance criteria, this also briefly covers performance assessment and training

effectiveness evaluation. The chapter also includes a detailed explanation of the representative complex cognitive task of helicopter overland navigation operations and current training.

Methodology: This section includes a description of the rationale behind the experimental design. This information is included to provide insight into design considerations. Coverage of the experimental protocol contains a detailed description of how data was collected.

Analysis and Results: This section covers novel data analysis techniques that were developed to both visualize and analyze the collected data. Results from this analysis are presented.

Discussion: In the discussion section there are details on the meaning and interpretation of our results and the relationship to existing literature.

Conclusion and Future Work: This section summarizes the major findings and contributions in the context of a larger study profile and suggests avenues for further exploration.

II. BACKGROUND

A. RESEARCH GOAL, SCOPE AND CONTEXT

This section provides a detailed view of the goals of the research (i.e., *what* we are trying to establish.) It also provides background information to explore the context and potential impact of this research (i.e., *how* this research fits into a larger conceptual framework.)

This research aims to improve the basis for providing individualized instruction. Figure 1 depicts the current basis for individualizing instruction. Figure 1 outlines the information flow from a trainee interacting with the training environment to an instructional system. The left side of Figure 1 depicts the trainee as an example of human information processing model. A more detailed description of the human information processing model is provided in Chapter II.C.1. The trainee functions by processing incoming stimuli, selecting a response and ultimately interacting with the environment via some overt response. Much of the processing and many of the critical elements of learning are internal to the trainee. However, the only information available to the instructor is the trainee's overt responses.

The right side of Figure 1 depicts the instructional system. This model of instruction is developed exclusively to provide context for this research. The instructional system compares a referent model of the trainee with the currently observed trainee state. The comparison of a referent model and observed state is used to drive the process for selecting and applying training interventions. Thus, the quality of instruction is limited by the accuracy of the instructional system's model of the trainee's state.

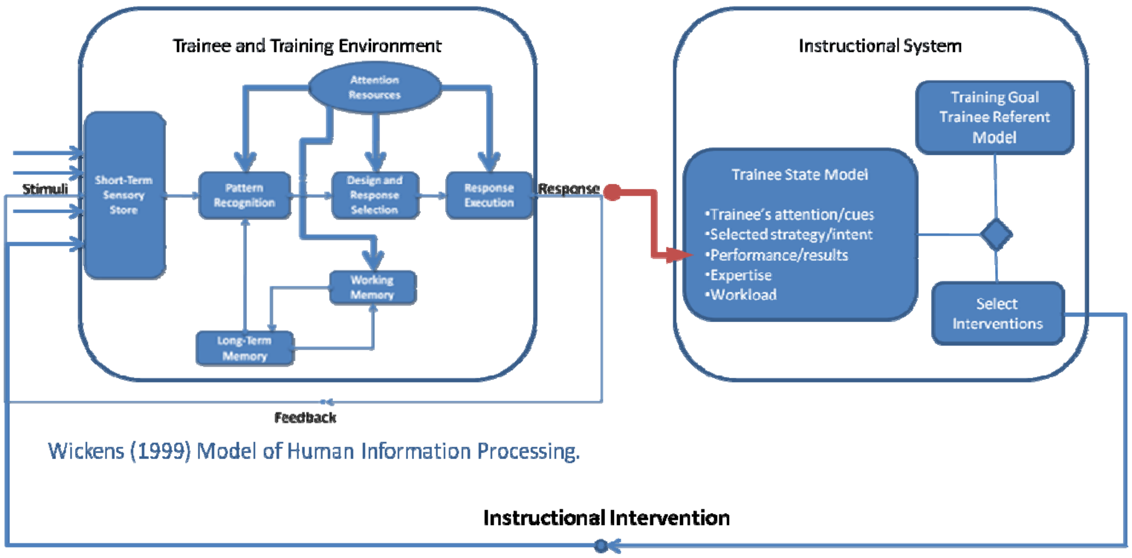

Figure 1. Current relationship of trainee and instructional system

Figure 2 depicts the general framework for improving the process for selecting training interventions. Quality of individualized instruction, as defined by selection and application of the most appropriate training interventions, is driven by the accuracy of the instructional system's model of the trainee's state. The research proposed here explores methods to select and validate neurophysiologic signals that can provide reliable indicators with diagnostic value to inform an instructional system of trainee state. By incorporating signals between the trainee and the instructional system that are not human-observable, we hope to provide improved awareness of trainee state. More detailed and reliable information on trainee state should ultimately improve the process of tailoring instruction, for example by enabling better selection of training interventions.

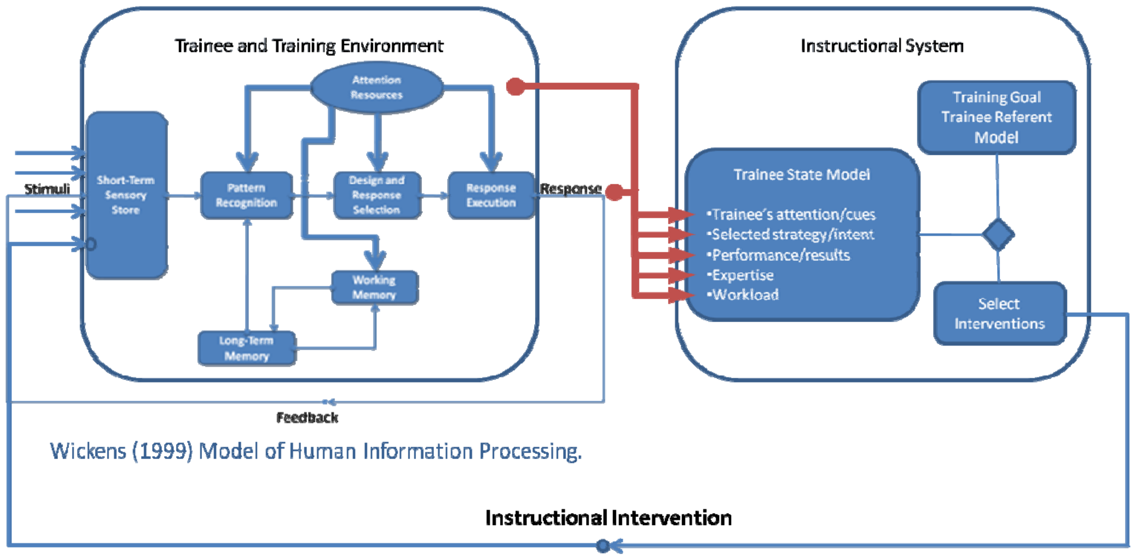

Figure 2. Improved information on trainee state

The accuracy, sensitivity and ease of use of neurophysiologic sensors will improve, as will our underlying models of human performance. This research explores how to use neurophysiologic sensors to derive reliable information related to an individual trainee's information processing. To be useful to a training system, neurophysiologic sensors must provide timely, reliable information with meaningful diagnostic value. The focus of this research is to build a model of novice and more-experienced pilot's visual scan patterns for helicopter overland navigation and validate this model with empirical data. Selection and application of instructional interventions is envisioned as a useful outcome of these efforts, but is beyond the scope of current research efforts.

Given the research goal of improving the reliability and usability of trainee neurophysiologic data to provide information on trainee's internal cognitive processes, the following section describes the larger context and overarching implications of this research effort.

The current stages for training design and evaluation of stand-alone simulation are depicted in Figure 3. Training requirements are developed and expressed in terms of resultant trainee performance proficiency. Training systems specification follows prescriptive methods that vary in their degree of connection with underlying cognitive

theory. If simulation is tailored to individual's level of expertise, it is based on classes of user's level of expertise. This level of individualization results in tiered simulation appropriate for groups of users. System implementation involves making tradeoffs against desired and affordable/deployable solutions. When systems are fielded operationally, it is often left to the users to provide oversight, instruction and mentoring. Finally, when trained individuals are evaluated, the primary means of evaluation is performance-based measures. The following section compares the current approach with potential improvements to simulation training design that takes advantage of neurophysiologic markers.

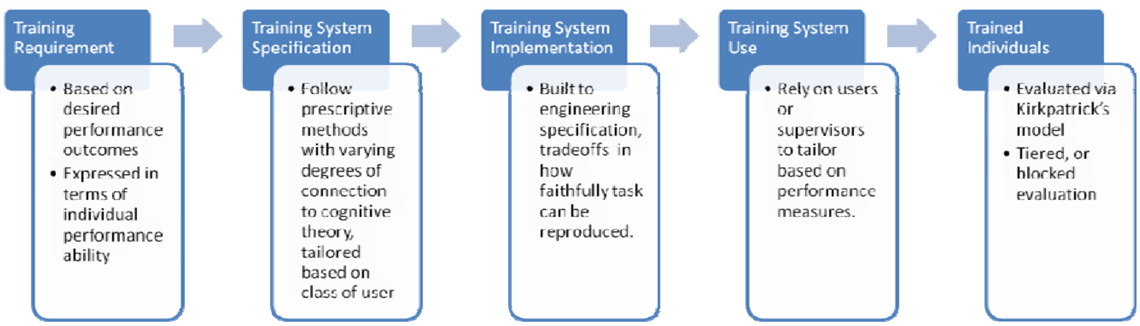

Figure 3. Current training and simulation development

Figure 4 depicts potential improvements to the process of simulation design and evaluation based on applying neurophysiologic markers. In the improved, augmented version of training simulation design, requirements would be specified based not just on raw performance measures, but on underlying user traits and measures that correlate to individualized level of proficiency. Training specification could describe appropriate simulation responses to key user's characteristics such as attention management, level of confidence and workload. Training could be implemented on tradeoffs in terms of impact on advancement in proficiency as indicated by internal state. System use would involve additional feedback on which an instructor or instructional system could base training intervention selection decisions. Finally, training evaluation could be based on both performance metrics and the underlying internal processes that contribute to improved performance, rather than relying exclusively on performance metrics.

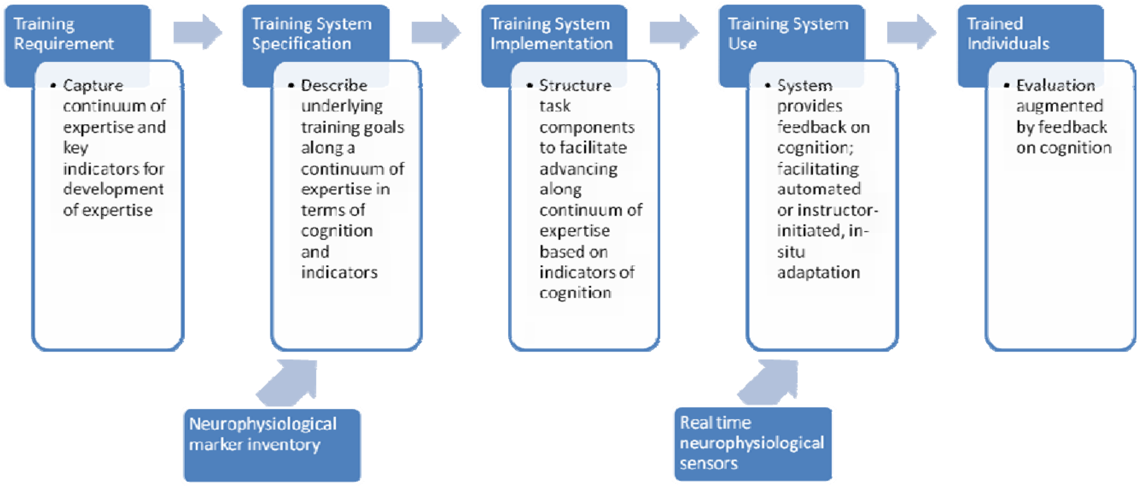

Figure 4. Implications of applying neurophysiologic markers on training simulation development

B. THE TWO-SIGMA PROBLEM

Over 25 years ago, Benjamin Bloom and three of his graduate students (Bloom, 1984) analyzed three forms of instruction and determined that individualized tutoring was vastly superior to alternative methods. They found that on average, students who received individualized tutoring scored two standard deviations above the control group that received conventional classroom training. The other treatment group applied Mastery Learning (ML) technique. Mastery learning took 30 years to develop and achieved a 1 standard deviation improvement; Figure 5. Bloom interpreted the results to suggest that any group of learners could achieve the higher levels of performance and defined this as 'The Two Sigma Problem.' Bloom (1984) further defined the objective and the challenge as follows: "I believe an important task of research and instruction is to seek ways of accomplishing this under more practical and realistic conditions than the one-to-one tutoring, which is too costly for most societies to bear on a large scale."

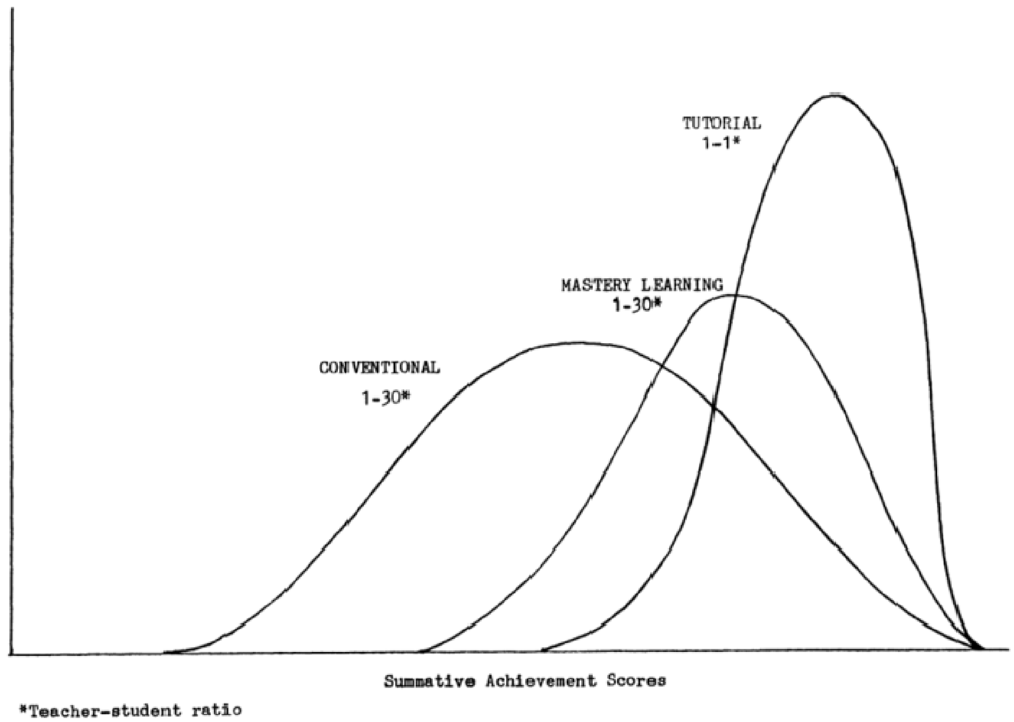

Figure 5. Test Scores for Conventional, Mastery Learning and one-on-one Tutoring. From (Bloom, 1984)

As noted by Bloom, unfortunately, individualized tutoring is not a practical solution for every training task and every group of learners. Providing individualized tutoring across all learning tasks, for all learners is neither feasible nor would it be cost effective. Since 1984, the variety of topics addressed, media employed, network infrastructure available, automation techniques and artificial intelligence methods in use have all expanded dramatically (Loftin, 2004.) And the two sigma problem still exists (Fletcher, 2004).

The story in the introductory chapter may provide some insight into viable areas for exploration. One critical question remains: what key element exists in individualized instruction provided by a tutor that doesn't exist elsewhere? Clearly, tutors rely on feedback from the trainee and are able to adjust their response in ways that most one-size-fits-all approaches and many automated systems fail to account for. For certain problem domains, such as tasks dominated by a psychomotor response, automated analysis of overt trainee actions such as physical motion may provide sufficient cues for

tailoring instruction based on an individual's traits. This approach has been demonstrated and provides early promise in individual tasks such as marksmanship (Platte & Powers, 2007) and some team tasks (Welch & Davis, 2008).

Unfortunately, it is not always possible to observe an individual's current execution or progress acquiring expertise in some tasks. In the opening example, even if an additional instructor could have been assigned to some imaginary position in the aircraft where he could observe the student, there would be no way to tell if the right sort of learning was taking place. Because the task relies on so many cognitive elements, there would be no way to determine how the student was performing on these task elements. For perceptual and cognitive tasks, many of the processes associated with performing and learning the task unobservable to the instructor's senses. Highly competent instructors may develop a sense of when a student is confident in their solution or when they are uncertain. Similarly, extremely seasoned instructors may develop a sense of when a student is overwhelmed; however, in the introductory story, it would be nearly impossible to tell what strategy the student may have selected and how successfully he correlated features from the out-the-window view with their contour map representation. The example demonstrates that there are cases where critical information to guide instruction is needed but is not readily observable. Part of the solution to the two sigma problem may lie in investigating signals that are not readily observed by humans but that can provide insight into a trainee's perception, attention management, level of certainty and workload while learning complex tasks.

C. CURRENT TRAINING-RELATED MODELS

1. Human Information Processing Model

The design, application and evaluation of training are based on assumptions about how humans process information. This section introduces a representative model of information processing to serve as a basis for explaining the hypothesis development and experimental design rationale as well as to providing a description of cognitive aspects of the selected training task and current training methods. Although information processing models provide a useful framework, they are not the primary focus of this work. While there are numerous models available in the literature, the sole requirement in selecting a

model is that it provides a solid basis for subsequent sections detailing research question definition and experimental design considerations. (Wickens, Gordon, & Liu, 2004) model of human information processing meets this criterion and is depicted in Figure 6. The Wickens model provides a basis for later arguments that allocation of attention resources is an important indicator of level of expertise and that passively detecting allocation of attention may be useful for guiding tailored instruction.

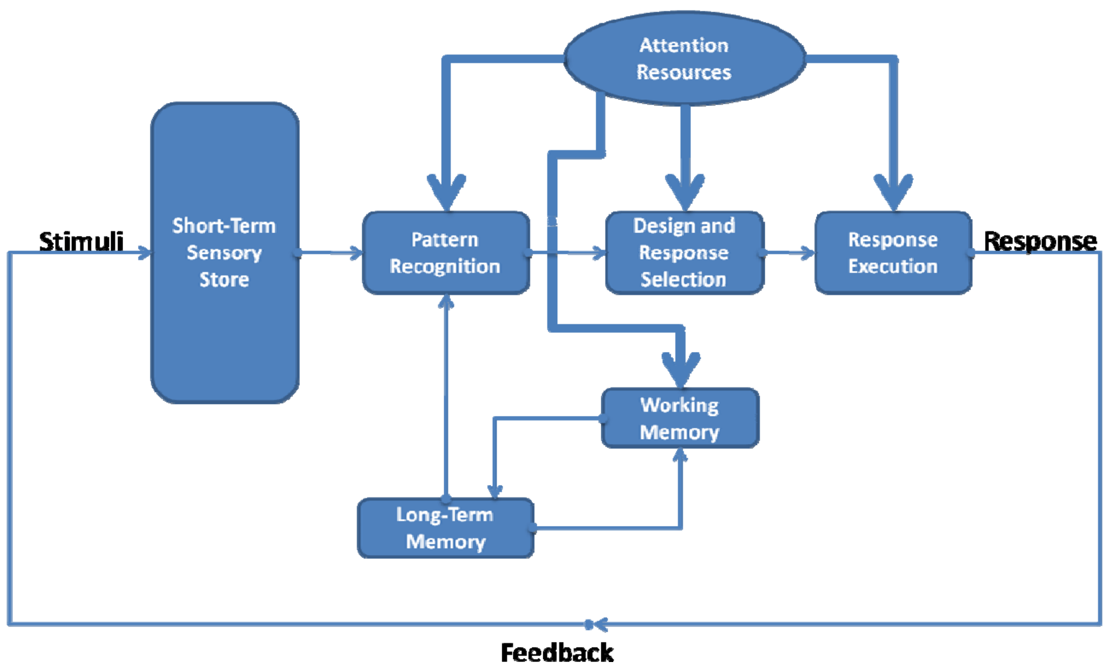

Figure 6. Wickens (1999) Model of Human Information Processing

Wicken's in (Welch & Davis, 2008) model assumes that information processing is divided in stages that form a continual loop. The loop is initiated with sensory processing. All sensory channels including visual and auditory have an associated short-term sensory store (STSS.) The STSS prolongs the representation of the raw stimulus. The duration varies from around on half second for visual STSS to around 2–4 seconds for auditory STSS. The next stage in the model is perception. Perception is the process of interpreting or giving meaning to the sensory data. Perception generally is an automated function requiring little attention. As depicted, perception is also affected by inputs from long term memory concerning events that are expected. Perception that is based on sensory data is referred to as bottom-up perceptual processing. In absence of

high quality sensory data, perception will be influenced by long-term memory and expectations; also referred to as top-down perceptual processing.

Cognitive operations are distinct from perceptual operations in that they generally require more time, effort or attention. Cognitive operations include mental rehearsal, planning, image transformation, and response selection. These operations are all conscious activities that transform or retain information, are resource limited and are highly vulnerable to disruption or interruption. For the most part, they operate with volatile working memory. They can also access more permanent, less vulnerable memory store referred to as long term memory. Encoding new information from working memory into long term memory is the basis of learning. The result of cognitive efforts can include selection and execution of a response. Selection of a response can involve a discrete event; however response execution often involves a continual and sometimes resource-intensive process.

The information processing model includes feedback and is affected by attention. Feedback is the process of observing changes in the environment based on operator input. One of the primary functions of a feedback loop is to determine if the operational performance goals have been achieved. There are two key aspects of this feedback loop that affect this study. First, the flow of feedback information can be initiated anywhere. For example, planning a flight path to avoid areas with ambiguous features and follow a prominent landmark will dramatically affect the quality of information available along the route. Thus, novices may inadvertently set themselves up for much more difficult tasks by failing to include prominent landmarks when planning navigation routes. Second, for tasks such as navigation and vehicle control, there is almost no delay in the feedback loop. An individual's overt actions can be instantaneous and not based on stimulus that is identifiable to an instructor. In the navigation example, both the definition and description of what constitutes a relevant feature can be very subjective and is particularly sensitive to individual's level of proficiency. The continuous nature and wide range of strategies that may be applied to complex tasks can make it difficult to assess performance. Across a broad range of tasks, instruction is difficult because evaluators lack cues that indicate subtle decisions that aren't associated with observable cues.

Attention is a key factor in Wicken's model. It involves mental operations that are not carried out automatically. Instead, they require the individual to selectively allocate limited sensory, perceptual or cognitive resources. Attention can be selectively allocated to various sensory channels and environmental cues. "For visual information, this limited resource is foveal vision, which can be (through eye movements) directed to different channels in the environment." By observing novices and experts scan patterns can we make inferences about the individual's sensory, perceptual or other cognitive factors associated with learning?

2. Complex Cognitive Tasks

The fundamental literature on human performance divides learning into three domains: Cognitive: mental skills (Knowledge); Affective: growth in feelings or emotional areas (Attitude); Psychomotor: manual or physical skills (Skills) (Bloom, 1984). Within the cognitive domain, various overlapping and interrelated terms have been used. (Schneider, 1985) addressed high-performance skills; defining the term, challenging assumptions for training programs and providing empirical characteristics of high performance skills. (Schneider, 1985) defined three characteristics for high performance skills. First, it takes a trainee considerable time (100 hours) and effort to acquire a high performance level. Second, a substantial part (greater than 20%) of a motivated population will not develop proficiency at the task. Third, there are substantial qualitative differences between novice and expert performance.

Van Merriënboer, Clark et al., (2002) adapts Schnieder's (1985) definition, further clarifies the characteristics of high performance skills and applies the term "complex cognitive skill" as equivalent. Where Schneider focused on a definition to support working guidelines to improve acquisition of skills, Merrienboer's definition was derived to support a novel Instructional Design (ID) model. This application is closely aligned with the intent of this research and thus will be adopted. (By other definitions, walking is a complex cognitive task). According to Merrienboer, complex cognitive skills have the following traits: They are made up of component skills; at least some of these skills involve conscious processing. The majority of the sub-skills are in the cognitive, rather than motor or affective, domain. According to this definition, examples include

computer programming, military air weapons control, air traffic control and many others. (J. van Merriënboer, Clark, & de Croock, 2002), (J. J. G. van Merriënboer, Kirschner, & Kester, 2003) further defines three characteristics regarding the structure of complex cognitive skills.

- They encompass a potentially large set of component skills that form a highly integrated hierarchy controlled by higher level strategies.
- Some sub-skills can be performed as automated processes while others are performed as controlled processes.
- They involve goal-directed problem solving skills. Expertise will be demonstrated in the degree of problem solving skills ranging from rule-based, scheme-based or weaker means-ends analysis techniques.

This definition and description of the structure of complex cognitive tasks provides a basis for future discussion of applying simulation to training. The nature of complex cognitive tasks—for example, the fact that they are comprised of a potentially large set of component skills involving a hierarchy controlled by higher level strategies—suggest the complexity of designing and evaluating instruction. Similarly, oversight of training implementation would be extremely challenging. Clearly, there is little useful information an instructor could gain from overt observation of task performance that would meaningfully inform an in situ or post-training event intervention process. Each of the bullet points above that comprise Merrienboer's definition of complex cognitive tasks provides an example of the information on trainee's internal state that would be useful for guiding instruction. Pausing to ask the trainee's thought process would alter the task to the point where it would no longer resemble an authentic task. Continuing without insight forces the instructor to rely on intuition and guesswork regarding processes internal to the trainee that impact learning. Complex cognitive tasks represent a unique subset of tasks that make it particularly appealing to investigate data streams that could be used to gain insight into trainee's cognitive processes.

3. Models of Expertise and Expert Performance

One of the shortfalls of current ISD models when applied to training complex cognitive tasks, is limited sensitivity to differences in task complexity, completion strategy and overall performance across trainee's levels of expertise (M. D. Merrill, 1990), (M. Merrill, 2002), (Patrick, 2004.) This section briefly reviews our understanding of expertise and previous efforts to map models of expertise to training methods. Current efforts are promising, but they have only been applied to one training domain and a single training medium. Additionally, they have yet to integrate automated methods for determining a trainee's current level of expertise, ability to perform component tasks, select high-level strategies or progress to the next level of expertise. Methods to assess performance and expertise at complex cognitive skills as cues to an instructional system are critical to fielding training systems that are reliable and available asynchronously. Training design and training effectiveness evaluation that is based on insight into factors such as a trainee's perception, attention management, level of task automaticity, confidence and workload could lead to more reliable design and evaluation when applying emerging technology to training complex cognitive tasks.

The goal of training systems is to create a measurable improvement in an individual's or a team's level of expertise. This section provides a brief review of the literature related to classification of expertise and reviews efforts to tailor training design and media selection based on an individual's levels of expertise. In the context of the Human Information Processing model presented in Chapter 0, both assessment of a trainee's current level of expertise and a trainer's selection of strategies for advancing expertise may be improved by searching for indicators of sensory, perceptual or cognitive management. Current assessment and intervention selection strategies are based primarily on overt performance characteristics.

a. Dreyfus & Dreyfus Five-Stage Model

The difficulties associated with capturing and describing expert performance date to the era of classic philosophy. Socrates observed that by definition, an individual adept at performing a task would also have lost appreciation for and ability to describe the nature of how the task was performed (Dreyfus & Dreyfus, 1984). In

1986, Dreyfus and colleagues proposed a five-stage model of expertise (Dreyfus, Dreyfus, & Athanasiou, 1986) . They postulated that any skill training procedure must be based on some model of skill acquisition so that it could address issues involved in facilitating advancement. This model has evolved considerably but is still relied on as the basis for much of the research related to expertise.

For each defined stage, Dreyfus and Dreyfus described the knowledge basis, mental effort involved and characteristics of performance. For example, the first stage or 'novice' level, is characterized by rule-based, effortful behavior that is limited and inflexible. Novices have learned, and rely primarily on, factual data. Because their knowledge is based on application of rules, they will have difficulty determining priorities of related tasks and adapting to evolving scenarios. In the next stage, 'Advanced Beginners' have gained and rely on domain experience. While their situational perception is limited, they can recognize meaningful aspects of situations. Advanced beginners have a level of competence; however they have difficulty determining priorities and can be easily overwhelmed. As expertise develops across the remaining stages, an individual's ability to recognize more abstract features of situations as well as the strength and reliability of their mental models improves. As an individual gains expertise, they become less reliant on rules and rigid guidelines. Their performance is less effortful and becomes more flexible and fluid. They respond in a more intuitive and automatic manner. Their description of their own mental models, processes and rationale may be incomplete or inaccurate.

Dreyfus' motivation for developing this taxonomy was to guide design of training. They provide the following high-level guidance (Dreyfus, Dreyfus, & Athanasiou, 1986) :

> *…The designer of training aids and courses must at all times be aware of the developmental stage of the student, so as to facilitate the trainee's advancement to the next stage, and to avoid the temptation to introduce intricate and sophisticated aids which, although they might improve performance at a particular level, would impede advancement to a higher stage, or even encourage regression to a lower one.*

The temptation is strong to apply technology exactly as Dreyfus' describes, and to conduct training effectiveness evaluation that measures the performance

gain while ignoring the underlying model of expertise. There are notable exceptions to the general trend to facilitate advances in performance while ignoring underlying development of expertise. Although these are limited in application domain (tactical decision making) and technology (personal computer simulation), the proposed framework is extremely promising. It highlights the significance and potential of systems that are aware of and responsive to a trainee's underlying level of expertise.

b. Ross, Phillips and Klein's Framework

Technology provides an appealing solution to training complex cognitive tasks. However, reliable guidelines for applying low-cost distributable simulations have not been identified. (Ross, Phillips, Klein, & Cohn, 2005) addressed these issues by developing a framework geared toward "…effective design, use and assessment of technology-based training for complex cognitive skills." This framework is geared toward Tactical Decision-Making Simulations (J. Phillips et al.) for training tactical thinking skills. The framework of (Ross, Phillips, Klein, & Cohn, 2005) is grounded in cognitive and instructional theory and provides a solid basis for research that will extend the literature on training design and evaluation. The framework has not been validated across domains. Testing across domains and exploring real-time methods to provide instructor's insight on a trainee's information processing ability and level of expertise could provide valuable guidelines for effective design, use and assessment when applying technology to train complex cognitive skills.

The general framework that (Ross, Phillips, Klein, & Cohn, 2005) developed consists of "…1) a five-stage model of learning and the characteristics of each stage; 2) the principles of the advanced learning essential to move from one stage to the next; and 3) implications for training each stage." The authors provide details for each stage of expertise – both general characteristics and characteristics specific to the domain of tactical thinking. The general characteristics include descriptions of both knowledge and performance. For each characteristic described, the authors provide a reference that extends and supports the original (Dreyfus, Dreyfus, & Athanasiou, 1986) taxonomy. For specific tactical thinking characteristics, the authors provide a description of the

tactical thinking profile and provide an example. These characteristics are compiled from previous studies and extensive work in the domain of tactical thinking.

Based on these descriptions of task characteristics the authors connect the general characteristics of trainee's with training and assessment implications. These training and assessment implications are supported by existing literature. A sample is listed in Table 1.

Stage 2: Advanced Beginner			
Knowledge	Performance	Training Implications	Assessment
Some domain experience (Benner, 1984; Dreyfus & Dreyfus, 1986) More objective, context-free facts than the novice and more sophisticated rules (Dreyfus & Dreyfus, 1986) …	Is marginally acceptable (Benner, 1984) Combines the use of objective, or context-free facts with situational elements (Dreyfus & Dreyfus, 1986) Ignores the differential importance of aspect of the situation; situation is a myriad of competing tasks, all with same priority (Benner, 1984; Dreyfus & Dreyfus, 1986; Shanteau, 1992) …	… Seeks guidance on task performance from context-rich sources (e.g. experienced people, documentation of past situations) rather than rule bases (e.g., textbooks) (Houldsworth et al., 1997). …	… Shows initial signs of being able to perceive meaningful patterns of information in the operation environment (Benner, 1984). Ignores the differential importance of aspect of the situation; situation is a myriads of competing tasks, all with same priority (Benner, 1984; Dreyfus & Dreyfus, 1986; Shanteau, 1992). …

Table 1 Stage Model of Skill Acquisition. From (Ross, Phillips, Klein, & Cohn, 2005)

The model by (Ross, Phillips, Klein, & Cohn, 2005) serves as a viable, although yet to be validated, method to design distributable simulation-based training that is sensitive to levels of expertise. Validating this framework could provide significant improvements over traditional ISD methods.

Ross and colleagues (Ross, Phillips, Klein, & Cohn, 2005) integrate a broad range of research and application into a cohesive, grounded and credible framework. The framework provides a promising alternative to current instructional design tenants to take advantage of distributable simulation for training complex cognitive tasks. The Ross model takes an important step toward realizing Dreyfus' aim for training. It appears to provide a viable means to focus on training design methods that facilitate advancement in the level of expertise as opposed to simply focusing on technology to achieve incremental improvement in performance. The (Ross, Phillips, Klein, & Cohn, 2005) model highlights a critical gap in existing literature related to expertise, simulation and complex cognitive tasks. It provides a framework for applying simulation; however we lack a means to diagnose a trainee's level of expertise as the basis for staged training, which prevents us from validating this framework. What is needed now is a more reliable means to make effective use of simulation technology to train complex cognitive skills.

c. Deliberate Practice

This section provides a brief overview of the concept of *deliberate practice* (Ericsson, Krampe, & Tesch-Römer, 1993). This discussion provides a general background on current research related to expertise and supports the classification of expert used in the study. This section also provides motivation for exploring methods to determine cues into cognition of novice versus expert and the potential to use these cues to improve training that advances expertise.

Deliberate practice is a theoretical framework supported by empirical research to explain how experts take advantage of tailored practice to achieve exceptional levels of performance (Ericsson, 1985) and (Ericsson, 1988) . The framework attempts to counter earlier notions that expertise was primarily based on heredity (Murray, 1989.) It also attempts to address why some individuals plateau and maintain sub-optimal stable

performance while others show continual improvement over as much as a 10-year period. The framework integrates some early investigation into the nature of expertise (Lesgold, 1983) and cognitive differences between novices and experts (Chase, Lyon, & Ericsson, 1979).

> Until quite recently researchers commonly believed that percentages of muscle fiber types and aerobic power "are more than 90% determined by heredity for males and females" (Brown & Mahoney, 1984, p. 609). Some researchers have therefore reasoned by analogy that basic general characteristics of the nervous system, such as speed of neural transmission and memory capacities, have a genetic origin and cannot be changed through training and practice

Countering earlier notions that cognitive ability was a predetermined characteristic, (Chase, Lyon, & Ericsson, 1979); (Chase & Ericsson, 1982) compared ability to recall chess board arrangements. They found that their ability was not based on predetermined characteristics, but was based on strategies and learned capabilities they developed for organizing information. They also investigated the domain-specific nature of expertise. They conclude that poor transfer of expertise across domains further supports the position that expertise is not based on innate characteristics but rather can be attained with the correct practice. The ultimate conclusion of their line of research is that practice plays a much more prominent role in the development of expertise than previously believed. They further characterized the nature of this practice and distinguished it simple repetition and rote execution. The nature of deliberate practice is that it is intense, effortful, focused on manipulating strategies and relies heavily and adjusts continually based on process feedback.

The origin of expertise has been studied extensively for decades. Early research (Trowbridge & Cason, 1932) demonstrated that mere rote repetition was not sufficient to ensure performance improvement. There also appear to be limits on the duration of practice that leads to meaningful improvement: "…deliberate practice is highly structured activity, the explicit goal of which is to improve performance. Specific tasks are invented to overcome weaknesses, and *performance is carefully monitored to provide cues* for ways to improve it further."

Deliberate practice impacts this research in two ways. In Chapter III.D.1, the definition of expertise for the experimental task is based on experiences that constitute deliberate practice. One of the critical components of deliberate practice relates to selection of strategy and awareness of results. For cognitive tasks, relying on observable performance cues may not adequately inform an individual or instructional system of the knowledge of results and task awareness required for deliberate practice. The limitations of observable cues and potential of integrating information that is not human-observable is central to the hypothesis developed in Chapter III.A. Specifically, the present dissertation is focused on providing feedback regarding trainee's cognition that could provide improved knowledge of results on cognitive tasks to inform an individual, instructional system or theoretical framework such as deliberate practice.

4. Cognitive Load Theory

Cognitive load theory (CLT) was derived by (Sweller, 1988) based on work of (Miller, 1956). Miller defined the cognitive limits as the 'magical number' of seven plus or minus two. That is, short term working memory could process five to nine items at one time. Later researchers observed that experts could effectively handle much greater capacity by apparently organizing information and procedures into 'chunks.' (Chase & Simon, 1982). Chase and Simon found that while underlying cognitive stores had similar capacity, experts organized information more efficiently and concluded that these skills were learned via repeated practice. The work of (Miller, 1956) and (Chase & Ericsson, 1982) provide useful insight into underlying cognitive models and expertise; however, their work was not directly useful for the design and evaluation of instruction. Sweller's work dramatically expanded the understanding of limits on working memory, mechanisms for developing schemas and implications for designing effective training systems. Drawing on the work of Chase and Simon, Sweller (1988) developed and tested theories on how to aid novices; for example by providing surrogates for the schemas that experts have developed and can use to reduce their cognitive load.

The main premise behind Sweller's work is that instruction should be designed around the limits of the trainee's cognitive load. Sweller (1993) define three types of load associated with instruction. Intrinsic cognitive load is the effort required to encode

new information; it is an inherent part of all instruction. Extrinsic cognitive load results from any mental tasks that distract from the main learning task and degrades learning. Poor arrangement of text, along with pictures, encoding of redundant information via multiple channels, and poor matching of sensory channel and learning content all increase extraneous cognitive load. Germane cognitive load is also external to learning the task itself, but unlike extraneous cognitive load, can be manipulated by the instructor to increase learning and transfer. For example, Sweller (1988) demonstrates that when learning trouble shooting procedures, randomized rather than blocked arrangement of problems leads to higher workload during learning, but improved far transfer during subsequent evaluation.

In addition to advancing the understanding the effect of cognitive load and media effects on learning, Sweller's work provides important insight on the differing impact of tailored training for novices and experts. Through a series of studies that varied the media and structure of training that is provided to novices and experts, Sweller discovered that techniques that can improve learning for novices may actually interfere with learning for experts. The nature of skilled memory (Chase & Ericsson, 1982) (Charness & Tuffiash, 2008), makes it more difficult for experts to learn using systems that work well for novices. In a counter-intuitive finding, (Chandler & Sweller, 1991) also discovered that experts can actually learn better from text that is intentionally made less coherent. Combined (Kalyuga, Ayres, Chandler, & Sweller, 2003) and (Sweller, 1988) terms the tendency of training methods that support improvements in novice to repress training in experts the expertise reversal effect.

Sweller's work (Sweller, 1988) takes important steps to connect underlying cognitive theory with practical design guidelines. The empirical results are an important guide for improved development of multimedia learning that accounts for novice versus expert differences. Unfortunately, cognitive load theory provides very little guidance for two important areas that would be necessary for developing similar guidelines for using simulation to training complex cognitive tasks: there are few guidelines and relatively little empirical research validating adaptive training, adaptive training is based on frame-based rather than highly dynamic first-person interactive simulation and there are currently no mechanisms for gaining insight into cognitive load during simulation events.

Thus, the ability to distinguish between novice and expert behavior will be critical for adaptive training and using simulation as a medium for complex cognitive tasks.

The discussion of CLT highlights some of the motivation for exploring cues of trainee cognitive factors such as workload as a means to provide knowledge of results to a trainee and to inform instructional systems. From the discussion in Chapter II.B.3.a, novice's performance is more effortful than experts. Based on theory discussed in Chapter II.B, this increased workload is likely because experts have developed more efficient structures for using memory stores. Insight into cognitive indicators such as the level of automaticity and ability to manage attention are ideally suited as a basis for customizing instruction demonstrated to be so much more efficient than a one-size-fits-all approach. Unfortunately, the current task analysis and training design processes discussed in Chapter II.C do not accommodate discovery and application of cues into cognition. The hypothesis developed in Chapter III.A takes initial steps in demonstrating the efficacy of such an approach—namely by validating that neuromarkers can provide insight into trainee's cognitive processes related to developing skill and thus provide meaningful diagnostic information to an instructional system.

5. Spatial Knowledge Acquisition

Subsequent sections provide a detailed description of helicopter overland navigation as an exemplar complex cognitive task useful for investigating neuromarkers that might provide useful insight into skill acquisition. This section provides background on spatial knowledge to provide context to support these arguments. Spatial knowledge is as a critical component of the larger task of helicopter navigation.

Spatial knowledge, or an individual's cognitive representation of a large-scale navigable space, is defined by (Siegel & White, 1975) as occurring in three phases. During an initial or landmark phase, knowledge consists of a set of disconnected landmarks. With increasing exposure to the environment, individuals learn to link important landmarks together. This level of knowledge is referred to as route knowledge. In the third level, referred to as configurational knowledge or survey representation, knowledge is represented in a flexible, map-like form. With survey level knowledge, individuals understand the spatial relationship of landmarks without relying on

knowledge of the routes that connect the landmarks. Individuals with survey level knowledge are better able to make spatial inferences independent of orientation. (Waller, Hunt, & Knapp, 1998) Survey level knowledge is an important component task of the overall task of helicopter navigation.

There is a direct analogy between levels of spatial knowledge acquisition and levels of task performance (Chase, 1982) (Chase & Chi, 1979). With landmark level knowledge, we could expect deliberate and effortful performance with little chance of adaptability and weak error recovery. With route level knowledge performance is less effortful. Individuals can rely on schemas and invoke scripts for completing routine elements of the task. They are not fully advanced as they may not be able to gracefully recover from errors and have limited ability to adapt to changes. With survey level knowledge, individual's performance would involve much less deliberate and conscious thought. A map-like understanding of the environment would allow them to recover from errors gracefully and easily adapted to changes.

D. CURRENT TRAINING DESIGN AND EVALUATION METHODS

The following section reviews current training design and evaluation methods: instructional systems design (ISD), cognitive task analysis (CTA), scaffolding, part task training, training effectiveness evaluation and simulation best practices methods and their related literature. In this section, we consider the role of neuromarkers in existing training literature. In the discussion section, we will refer back to how these processes might be improved for complex cognitive tasks by adding elements of neuropsychological feedback mechanisms.

1. Training Design (ISD)

Instructional systems design is an engineering approach that defines a process for creating courseware, curricula and learning media for education and training. The precursors to instructional systems design were originally developed around World War II. Original models (Briggs & Ackerman, 1977) were developed by the United States military to help standardize the process for creating instruction in an effort to dramatically improve the efficiency and bandwidth of military training programs. To

varying degrees, ISD provides prescriptive measures that connect underlying cognitive principles and theories with practical implementation guidelines (Dick, Carey, & Carey, 2007). Since its early definition and implementation, ISD has evolved considerably in response to advancing cognitive theories, varieties of training tasks and available technology. While ISD has dramatically improved development of curricula and supporting media, it is not well suited for defining the role of simulation in the acquisition of complex cognitive tasks.

Figure 7 (Dick, Carey, & Carey, 2007) depict a popular and representative ISD model. Although the model appears linear, the authors stress that the model is intended to be applied in an iterative manner with frequent deviation as individual training problems and practitioner experience and preference dictate.

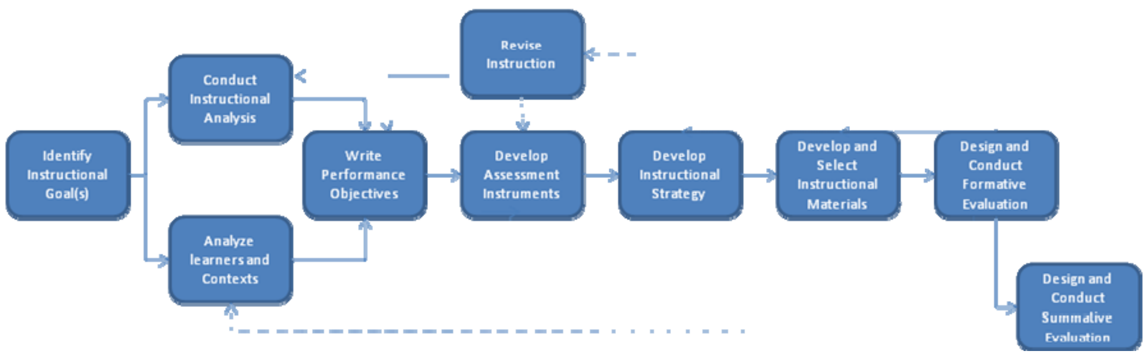

Figure 7. The Dick, Carey Model of Instructional Systems Design. From (Dick, Carey, & Carey, 2007)

Reviewing the application of this model in chronological order illuminates strengths as well as potential limitations of applying ISD for the design of simulation for training. ISD models generally include a front-end analysis phase. One main purpose of the front end analysis is to ensure that there is a well defined problem for which training is a likely solution. In some cases, poor performance is inappropriately attributed to a training issue. Often other factors, such as inadequate job performance aids or poor human factors design are the root cause. In these cases, there may be straightforward solutions that are easier and more appropriate than training. The outcome of this stage is: "…(1) a clear, general statement of learner outcomes that is (2) related to an identified problem and needs assessment, and (3) achievable through instruction rather than some

more efficient means…"(Dick, Carey, & Carey, 2007)   An inherent strength of any systems approach and specifically of ISD is a clear definition of the problem. The list of goals (capabilities of learners) developed in this stage is also useful later for defining assessment of learning. Listing goals is a major strength of the ISD process but does not necessarily mean ISD is appropriate for clarifying the role of simulation in training.

The next phase involves two stages depicted in parallel: Conduct Instructional Analysis and Analyze Learners and Context. The instructional analysis phase is divided into two major steps: Goal Analysis and Identifying Subordinate Skills and Entry Behaviors. "The main purpose of the goal analysis is to provide an unambiguous description of exactly what the learner will be doing when performing the goal."

For complex cognitive tasks, there are several potential issues with this stage. An 'unambiguous description of exactly what the learner will be doing' may not be practical to achieve, useful for training design or reusable across development cycles. By definition, experts apply a wide variety of strategies and techniques when addressing complex cognitive tasks. 'What the learner will be doing' may cover a very diverse set of options, strategies and approaches (J. van Merriënboer, Clark, & de Croock, 2002) (J. J. G. van Merriënboer, Kirschner, & Kester, 2003). Task analysis requires an interdisciplinary approach with task analysis experts working with subject matter experts. This process can be subjective and lead to different products (Kirwan & Ainsworth, 1992). Defining characteristics of the desired end state for learners—expert performance—include the fact that there are multiple approaches available and experts will have difficulty articulating their approach (Ericsson & Lehmann, 1996). Further, describing the stages that novices progress through to achieve this expert performance is notoriously difficult (Ericsson, 2006). A detailed description of how an expert defines a task will be dramatically different from how to compose part-task training solutions into an effective training continuum. Although ISD is based on an understanding of cognitive science, there are gaps where the procedures of ISD may not match underlying cognitive science. For example, cognitive science theorizes that, as learners progress, they become more efficient at managing workload. 'Chunking' (Wickens, Gordon, & Liu, 2004) and automaticity both contribute to improved performance. These processes are critical to expert performance, but are not easily observed or described and may not be discovered

in task analyses. An expert's view and description of a problem will be significantly different than a novice's. For example, when reviewing computer programs, novices categorize programs based on surface features where experts have the ability to look beyond surface features and categorize programs according to their underlying structure. Thus, an expert's description of a task will be based on a different set of cues than a novice's. These distinctions may be subtle, difficult to capture, and can lead to substantial differences in task analyses.

The first step in the goal analysis process is to categorize the goal according to (Gagné, 2005) domains of learning. These domains are: verbal information, intellectual skills, psychomotor skills and cognitive strategies. The classification is intended to help align training goals with appropriate media. There are two significant issues. First, separating verbal information from its context can lead to inert knowledge (Whitehead, 1929). Learners who memorize facts and procedures without context or practice have demonstrated difficulty recalling information while executing the task. Second, this task decomposition scheme does not support constructivist learning theories. There is evidence that providing as nearly-complete a version of the task as early as possible in a trainee's experience leads to more effective learning and better transfer. (R. Clark, 2003). Breaking down a task based primarily on the type of learning involved does not tend to preserve key information required to re-assemble task components into logical part-task versions that lead to full-complexity versions of the task early in a trainee's experience.

There also are issues with the second stage of the Conduct Instructional Analysis phase: Identifying Subordinate Skills and Entry Behaviors. For each goal identified, this step involves breaking down the goal into sub-goals and prerequisite knowledge. The main difficulty is simply stated: "It is almost impossible to know when an appropriate and valid hierarchical analysis of an instructional goal has been achieved" (Dick, Carey, & Carey, 2007). The problem of not knowing how far to decompose a task complicates the process and increases the level of effort required to analyze goals. Lack of clear guidelines in decomposing tasks also makes it difficult to reuse task decomposition on subsequent training design cycles. If tasks are broken down too far makes it can be difficult to recombine the sub-tasks into meaningful practice events. Inadequate

decomposition that is not breaking the task down far enough, could lead to hidden prerequisite knowledge or omission of key sub-goals; subsequently leading to ineffective training systems.

The ***Analyze Learners and Contexts*** phase involves an in-depth look at the environment and trainees. The primary characteristic of learners considered is prior education level. Education level tends to overlook an important aspect of training systems used for complex cognitive skills. ISD tends to focus on training for novices. In contrast, complex tasks may take years to attain proficiency or expertise. It is extremely difficult to imagine that the same simulation training system used for training novices would be useful for training an advanced expert. ISD tends to focus on a single set of users and targets a single proficiency level.

The majority of the decisions related to training system design take place in the Develop Instructional Strategy phase. In defining instructional strategy, (Dick, Carey, & Carey, 2007) describe the relationship of micro and macro-strategies. Micro-strategies "… include a wide variety of teaching/learning activities such as group discussion, independent reading, case studies, lectures, computer simulations, cooperative group projects, and so on." Macro-strategies describe how each of these components will be used to facilitate the learning process. The focus on this phase of ISD is on macro-strategies. Instructional strategy covers "… choosing a delivery system, sequencing and grouping clusters of content, describing learning components that will be included in the instruction, specifying how students will be grouped during instruction, establishing lesson structures, and selecting media for delivering instruction." In this phase of ISD there are several potential shortfalls for describing simulation for training complex cognitive skills.

Dick and Carey (2007) define a delivery system as the methodology for "… managing and delivering the teaching and learning activities that we call instruction." Examples of delivery systems include the traditional model of an instructor in a classroom or lab, web-based instruction, and computer-based fully interactive multimedia instruction. Dick and Carey (2007) describe an ideal path for choosing a delivery system.

This path follows a sound systems engineering approach that is "… based on careful consideration of needs and requirements before a solution is named." They also point out a major difficulty with this phase of ISD.

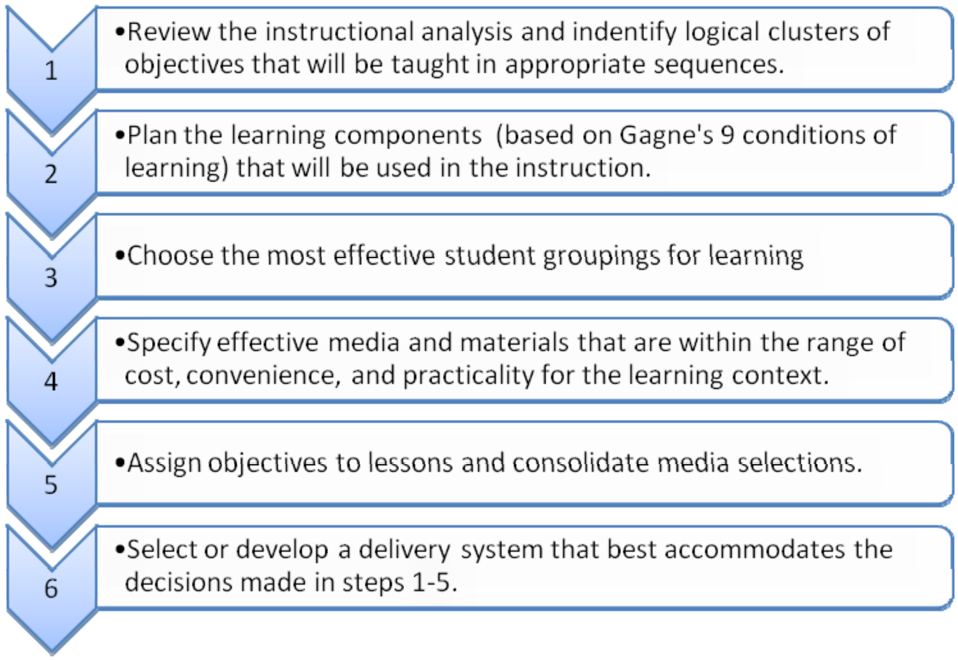

Figure 8. Dick and Carey's Instructional Design Process. From (Dick, Carey, & Carey, 2007)

This ideal path "…almost never happens." For a variety of reasons, the delivery system and its constraints are often selected before the requirements are fully known. In the context of DoD simulations, end users are usually more familiar with current and emerging technology than they are with a systematic approach to requirements definition. End users can inadvertently circumvent an otherwise sound systems engineering approach. The DoD acquisition process exacerbates this problem. Training acquisition programs are specified based on physical characteristics rather than desired end state such as training effectiveness or improved readiness. To paraphrase one systems acquisition professional: "We're called Naval Air Systems Command for a reason. We acquire training systems not training effectiveness." (Patrey, 2005)

Another issue with the ISD process is that training curricula evolve over time independent of the ISD process and can involve multiple independent systems. For many training domains, it may be difficult to catalog the capabilities of existing individual

training mechanisms. Fitting a new system into a cohesive continua composed of these disparate parts can be difficult. Deciding where a newly developed training capability should be integrated into existing curricula is not an inherent strength of existing ISD processes. In the DoD context, this can be problematic. Deployment schedules and constraints artificially limit availability and access to optimal training systems. DoD training systems demand a degree of flexibility both in the design of individual training systems and in the construction of cohesive training continua that is well beyond the capabilities of current training design methods.

Assuming designers were given the opportunity to follow the ideal path, there are still significant issues with designing simulation for training complex cognitive tasks. As previously discussed, it can be difficult to establish an appropriate hierarchy of goals and sub-goals. One of the steps in this phase of ISD involves clustering instruction. According to (Dick, Carey, & Carey, 2007), "The decision to cluster the objective is a subjective one… made based on knowledge of both the content and the learners." It is difficult to gauge how sets of learning objectives should be grouped to facilitate learning across a range of experience. Novices will require a simplified version of the task for initial exposure. Experts will benefit from a much different version of the task. Oversimplification of tasks can lead to formation of weak mental models and delayed learning. ISD does not provide guidelines regarding what these different versions should look like and how individuals should transition from one version to the next.

Instruction systems design's approach to media selection presents additional issues. There has been considerable discussion about the effect of media selection on learning. The seminal review by (R. E. Clark, 1983) and since backed up by (Russell, 1993) and Clark and Russell (1999) support the position "… that it is the design of instruction rather than the medium used to deliver instruction that determines student learning." This position could be affected by the definition of multimedia and by the difficulties comparing successive generations of multimedia capabilities. The capabilities of multimedia displays in 1983 have certainly evolved considerably over the last 25 years. It is difficult to imagine that selecting a 1983 version of a multimedia device would have the equivalent ability to deliver training and as cost-effective as a 2008, or 2033 multimedia device. In practice, describing and capitalizing on the potential benefits

of successive generation of ‘multimedia’ is difficult. The current range of multimedia presentation capabilities, including desktop personal computer simulation, is much more diverse, capable and inexpensive than current ISD methods provide insight for exploiting. In general, categorizing the capabilities of emerging technology into the ISD process is not well supported. Considering current personal computer gaming capabilities, a much different range of skills could be trained in a real-time strategy game than could be trained in a first person shooter. However, if designers view all media selection options as nearly equivalent, it is impossible to make such distinctions and capitalize on unique capabilities of various multimedia capabilities.

Even with more conventional media selection options such as graphics and audio, connection with the underlying cognitive science can be difficult. According to (R. Clark, Nguyen, & Sweller, 2005) “It is widely believed among multimedia instructional designers that duplicating informationally identical audio and visual material facilitates learning.” Their research indicates the opposite is true. Redundant encoding can actually interfere with learning. Additionally there are significant differences in learning for experts and novices. Kalyuga in (Kalyuga, Ayres, Chandler, & Sweller, 2003), describes the expertise reversal effect in which instructional environments that are effective for novices actually depress learning outcomes for experienced learners.

Although ISD methodologies are proven, reliable and have advanced considerably, review of each stage in the ISD process indicates areas where it is not well suited for designing training for complex cognitive skills. The hierarchical nature of component skills and variability from novice to expert performance make it extremely difficult to determine how to decompose the overall task, restructure it according to how skills develop and align these composed skills with available media for replicating the task. Design of training for complex cognitive skills requires improved methods for understanding how expertise is developed and evaluated.

2. Cognitive Task Analysis (CTA)

This section provides a brief summary of cognitive task analyses in the context of human information process models, expertise, instructional design and customized instruction. Current CTA methods do not readily provide a means for identifying

neurophysiological cues that could be useful for guiding instructional intervention. Although expanding CTA techniques to address finding neuromarkers to cue instruction is beyond the scope of this work, this section introduces a relevant CTA of a complex cognitive task. A review of this CTA, discussed in detail in Chapter 0, suggests that the CTA process may be useful for discovering or validating neuromarkers and key elements of cognition.

In their coverage of CTA, (Kirwan & Ainsworth, 1992) define task analysis as "…the study of what an operator (or team of operators) is required to do, in terms of actions and/or cognitive processes, to achieve a system goal." According to (Kirwan & Ainsworth, 1992) the purpose of task analysis is "…to compare the demands of the system on the operator with the capabilities of the operator, and if necessary to alter those demands, thereby reducing error and achieving successful performance." They identify six major human factors issues that influence system success; consider where in the system lifecycle these issues are best addressed, and outline analysis techniques that are most appropriate. They describe the human factors issues associated with the design of training as follows: "Determine decomposition level required to produce trainable units for effective learning (i.e., to ensure there is sufficient detail.) Determine how skills are best acquired, and whether an on-the-job instructor would have control over range of tasks, situations and events that would have to be dealt with. Is there a useful knowledge base that people would need? Is simulation required? Identify training methods." The authors note that there is a strong connection between the goals of a task analysis for training and task analysis for personnel selection. There are at least two critical issues. How do we account for individual differences that might change the definition and thus the task decomposition level required to produce trainable units for effective learning? Second, are there opportunities in the task analysis process to define the information available to an instructor or instructional system that would indicate when it is appropriate to assemble trainable units into more complete tasks? A task analysis procedure that could identify cues for assessing the effectiveness of defining and assembling trainable units could serve the training design process better. The methods Chapter 0 defines preliminary steps required to achieve this: validating that neuromarkers

identified in a CTA can distinguish between novice and expert behavior where the distinction cannot be made using overt behavioral observation.

(Schraagen, Chipman, & Shalin, 2000) cover the current state and future challenges for advancing CTA methods (Schraagen, Chipman, & Shalin, 2000). In this review (Lesgold, Feuer, & Black) recommends a future direction for CTA: "Perhaps we should be paying some attention to the collection of novice competencies that provide a potential pathway to expertise." As a leading example, Lesgold refers to the work of (Barnard) who describe task analysis as a collection of models of different levels of knowledge and performance capabilities. While Barnard and May provide a strong connection between their CTA methodology and underlying cognitive theory, their methods do not suggest any techniques for informing a training design process of what insights into an individual's cognition might be useful for tailoring instruction.

The recommendation for future work in CTA development is echoed in a more recent update (Crandall, Klein, & Hoffman, 2006). In particular, they point out the importance of accounting for differences in expert and novice performance (Hoffman & Palermo, 1991), (Ross, Phillips, Klein, & Cohn, 2005) for designing effective training systems. According to the model of cognition (Klein, 2008) , key differences between novices and experts are reflected in macrocognitive processes (Ross, Phillips, Klein, & Cohn, 2005). Further, attention management is central to effective macrocognitive function. Thus, current CTA literature supports the premise that cues that indicate novice and expert difference in attention management may be useful for designing tailored instruction.

3. Scaffolding and Part-Task Training

Part task training and scaffolding are recognized as valid mechanisms for improving instruction. Simplification via scaffolding or by breaking the compete task into more manageable sections has been proven effective when the compete version of the task can be too overwhelming for novices (Wainess, 2003). Unfortunately, the set up cues and information on how far to break a task down and when to assemble the parts into a more complete version is very limited. (R. Clark, 2003) points out several examples where breaking the task down too far actually interferes with learning.

Investigation of neuromarkers could provide valuable cues both into how far to break a task down and if applied as part of adaptive training, when to recompose parts of training towards a more complex whole task

Knowledge of a trainee's internal processes related to development of expertise could improve an instructor's ability to take advantage of part task training or provide scaffolding. With highly cognitive tasks, cues to estimate ease and accuracy of task execution may not be readily apparent. For example, a command and control task may involve directing aircraft via a variety of control measures (Schiaffino, 2005.) If certain control measures are preferable but require more effort, novices may easily be overwhelmed. They may be drawn to use control measures that are sub-optimal, but easier to execute. If an instructor could recognize that the effort involved for a trainee to execute preferred control measures was excessive, he may tailor instruction differently. He could structure training opportunities to allow the student to practice that control measure until it could be performed as an automated script. Of course, this adaptive process is exceedingly difficult since we cannot currently sense the trainee's level of difficulty for the overall task or level of automaticity for component tasks.

4. Training Effectiveness Evaluation

Patrick (1991) presents a concise overview of training effectiveness evaluation, relying on Goldstein's (1986) definition: "…the systematic collection of descriptive and judgmental information necessary to make effective training decisions related to the selection, adoption, value and modification of various instructional activities." Evaluation of training normally follows the four levels defined by Kirkpatrick (1967.) Reactions measure user acceptance and credibility. Learning, generally conducted using post-tests, evaluates the degree to which knowledge was effectively imparted. Job behavior examines performance differences. Results measure the actual impact on the organization: did any improvement in the trainee's job performance actually improve the overall ability of the organization to conduct its mission? Despite the seeming importance of training effectiveness evaluation, TEE occurs rarely. Phillips (1990)

reported on a survey of management training. In this report 52% relied on trainee feedback, 24% measured a change in job performance and less than 2% measured any return on training investment.

Since 1990, there have been significant if not tremendous advances in our ability to create virtual environments for training. Yet, as is frequently pointed out in (Schmorrow, Cohn, & Nicholson, 2009) the frequency of TEE and the science of both formative and summative training evaluation has not advanced at the same pace as the underlying media for training. Identifying neruomarkers that could be used to indicate differences in novice and expert performance and provide insight into a trainee's cognition could contribute to a reliable means to both guide training design and also evaluate the resulting training system. Evaluating performance alone does not always indicate improvements (R. Clark, 2003). Neuromarkers could provide an important connection between how we describe tasks, define training design goals, and assess trainees and training systems.

E. TASK AND NEUROPHYSIOLOGICAL MARKER SELECTION

The purpose of this section is to provide background on helicopter navigation as a representative complex cognitive task for studying the potential for neurophysiologic measures to be used as a guide for individualizing instruction. The description of helicopter navigation and expertise supports the rational and process for selecting eye scan as an appropriate measure. Based on the cognitive task analysis we build a model of novice and experts scan that will serve as the basis for our experimental design.

1. Helicopter Overland Navigation

Helicopter overland navigation meets the criteria described previously as a complex cognitive task. This section discusses the task in detail and provides background information to support experimental design rationale. This discussion is supported by a cognitive task analysis originally developed in (Sullivan, 1998) and further refined by (McLean, 1999), (Lennerton, 2004), (Beilstein, 2003), (Kulakowski, 2004) and (Hahn, 2005). The full task analysis is included in Appendix A.

a. Task Description

Helicopter navigation relies on a number of sub-skills. These sub-skills are highly interdependent and form a hierarchy that requires high-level selection of strategy for proper execution. Several of these sub-skills can be performed as automated processes while others required controlled processes. Finally, this skill involves goal-directed performance in which trainee's level of expertise is indicated by their goal-seeking method: rule-based, schema-based or means-ends analysis.

Navigation is normally the role of the non-flying pilot. The non-flying pilot is responsible for providing verbal instruction to the flying pilot to reach navigation check points. As described in (Wright, 2000), navigation is never the sole aim of a mission. It a necessary goal required for completing a higher level task such as logistics support; intelligence, surveillance and reconnaissance, or combat search and rescue. The non-flying pilot will have additional responsibilities including terrain and obstacle avoidance, monitoring and managing engine and system performance, and communications. In this respect, navigation is sub-skill of a larger complex cognitive skill that is beyond the scope of this study.

Pilots rely on a variety of techniques for overland navigation. Each of these techniques can be considered as a sub-skill following the definition of complex cognitive skills. Consistent with the previous description of the structure of complex cognitive tasks, these sub-skills form a hierarchy, involve both automated and controlled processes, and expertise is demonstrated in the level of goal-directed behavior. One of the techniques involved is dead reckoning. Dead reckoning involves using ground track, ground speed and timing information to estimate current position as a function of previously known location. The component skills for dead reckoning can be practiced to a level of automaticity. Recovery from dead reckoning errors is difficult if not impossible, thus it is only relied on as the sole method of navigation if there are no other options such as navigating featureless desert at night.

Terrain association is the process of identifying unique features or combinations of features from the out-the-window view with the corresponding two-dimensional representation on a contour map. This sub-skill represents a controlled

process that is labor intensive but more robust. Component skills within this task include determining aircraft heading, estimating position by triangulating between terrain features and communicating with crew members who may have more salient terrain features within their field of view. Recovery from terrain association errors can be much more reliable than from errors while dead reckoning.

Normally pilots will employ some combination of techniques to maintain a reliable navigation solution while attempting to minimize workload and intra-aircraft communication. Pilots employ high level strategies to determine which technique to employ based on currently available cues. High level strategies also are used to evaluate the effectiveness of the combination of selected methods.

Novice pilots typically demonstrate means-ends analysis when approaching navigation tasks. More experienced pilots will demonstrate schema-based behavior and experts will use rule-based procedures. For example, a novice pilot will tend to rely heavily on dead reckoning skills regardless of terrain features. Their selection of a combination of techniques will be labor-intensive and consume scarce cognitive resources. The tendency is to view the navigation course as a sequence of unrelated individual legs while ignoring opportunities to make gross simplifications to the overall task. Task simplification would include altering the flight path to conform to prominent features that nearly parallel the intended flight path. Since the process of map study – that is the selection of prominent distinguishing features for use in navigation – depend on nascent, emerging parallel skill of feature identification, trainee's will typically attempt to identify too many potentially helpful cues. Novice navigator's general strategy will reflect linear processing of successive landmark cues typical of means-ends analysis.

Intermediate navigators have typically developed some of the requisite sub-skills to a certain level of automaticity. This level of skill enables increased flexibility in selecting and executing strategies. Intermediate navigators will also have an improved ability to execute a collection of sub-tasks as a schema. For example they will be able to identify guiding features to use to restrict lateral deviation from the intended course. Their commands to the flying pilot will be based on fewer, more prominent

features. Typical schema execution will be reflected in the communications to the flying pilot. Navigation instructions will be more general-purpose and will demonstrate planning beyond the immediate field of view. For example, an intermediate navigator may direct the flying pilot to follow a guiding feature until encountering a certain prominent limiting feature and provide directions for initial turn and a description of the next prominent feature to look for.

Expert navigators will have achieved a high level of automaticity for the overall task and component skills. Their performance can be described as rule-based. They will have reduced the apparent complexity of the task to the point where they execute the overall task as a set of relatively straight-forward steps. For example, an expert may describe an entire navigation route in terms of several critical steps based on a coarse description of global navigation features. The rules they describe will reflect sound selection of navigation technique and robust backup techniques.

b. Training Environment Description

As is the case with many critical complex cognitive tasks, the environment for conducting helicopter overland navigation is not conducive to training. Physical layout of the cockpit, restrictions on route selection, safety of flight concerns—including aerodynamic constraints and communications restrictions—make it difficult for instructors to provide quality training.

The time, cost and risk associated with any flight operations are well known. As difficult as it is to allocate flight time to train navigation skills, the flight environment adds additional constraints. Selection of general terrain type, foliage and seasonal and weather effects – to include night vision lighting conditions – is beyond the control of the instructor. Navigation routes must be planned around population centers, noise abatement areas, civil traffic patterns, wildlife protection areas and numerous other constraints. Navigation training operations are typically restricted to areas and routes which very quickly become familiar to operators. Navigation tasks are reduced to recognition of prominent features identified during the most recent flight.

Once navigation flights are underway, there are still many features of the operating environment that make it difficult to train. Primary among these is crew workload. The flight profile is designed to minimize exposure to potential enemies. Flight paths are therefore planned to use low altitude and high speed to achieve terrain masking. At low altitude and high speed, it is vital that each crewmember maintains a vigilant scan pattern for terrain and obstacle avoidance. The non-flying pilot also is responsible for many other flight maintenance tasks including monitoring and managing engine and performance indicators and handling communications tasks.

The instructor faces numerous challenges to provide instructional opportunities. Cockpit layout makes it difficult to simultaneously control the aircraft, ensure safety of flight and monitor the trainee. Consider what should be the relatively straight-forward task of determining if a novice has correctly correlated a terrain feature within view with its representation on the map. Simply pointing to the feature within view is not feasible. Cockpit layout and flight profile make it impossible to disambiguate a terrain feature by pointing. The instructor and trainee rely on verbal exchange to mutually agree on a description of the visible feature. The instructor must then determine if the feature correlates to the contour map representation the student has identified on the map. This can be extremely difficult. Because the pilots must always maintain a vigilant scan ahead of the aircraft it is extremely dangerous for both pilots to focus inside the cockpit at the map at the same time. It is not practical or safe for the flying pilot to take his hands off the flight controls for long enough to perform a simple instructional task such as pointing to the correct location on the map. Again, the pilot must rely on verbal exchange to describe the correct map location.

The bandwidth is extremely limited for verbal exchanges to identify a unique feature within the field of view, the map representation of that feature and provide any instructional comments that may help the student gain expertise. All crew members are connected on a single interphone communications system (ICS) channel. This single channel is used for exchanging critical and timely safety of flight information. Extended conversations are not feasible. Instruction must be condensed into terse statements that allow for frequent pauses so crewmembers can inject required information. In a busy environment, for example when following a narrow canyon and relying on all

crewmembers to provide terrain proximity information, instruction to a disoriented pilot would be limited to phrases such as "peak 271 is the high peak at your 10 o'clock." In a more permissive environment, the instructor could feed cues to help orient the student and improve their skill such as "the hill at your 2 o'clock is uniquely identifiable by its conical shape and steep face oriented north to south."

The instructor is limited in choice of flight profiles that may support learning. Given the task complexity and inherent risk associated with terrain flight profiles, slowing the aircraft would seem like an appealing option. Unfortunately, slow flight at terrain flight altitudes increases flying pilot's workload significantly and substantially reduces the margin of safety. Because of the high power required and low usable potential energy in this flight regime, responses to any aircraft emergency is extremely limited. Slowing to allow a novice more time to process information or reorient himself is not practical. Options for reorienting are very limited. It is contrary to tactical doctrine to climb to achieve an increased visual horizon to see more features from a perspective where they more nearly match their map representation. Not only is this a bad habit to train, but terrain association at altitude is a dramatically different skills set than at terrain flight altitudes.

Other options for guiding student pilots are limited. Novice instructors may tend to consider circling a familiar landmark as a viable option for reorienting a student. In practice, this is extremely time consuming and disorienting. Much of terrain association depends on feature orientation relative to the aircraft's heading. When heading is constantly changing – even at a predictable rate – the complexity of scanning inside to determine heading, outside to determine terrain feature orientation, and back to the map to find correlation is overwhelming even to experienced pilots. Given the difficulties associated with re-orienting a student and the restrictions on route selection, instructors provide very narrow margins for error. Limiting the degree of errors students are allowed to make restricts the overall instruction that experienced pilots can provide.

Based on this summary description of the task and operational environment, helicopter overland navigation represents a complex cognitive task in an operational environment that is not conducive to training. It is based on components

skills. Some of these skills, such as dead reckoning techniques, can be automated. Other component skills, such as terrain association, are performed as controlled processes. The sub-skills form a hierarchy that is controlled by high-level processes. Expertise is demonstrated in the level of goal-directed behavior. Training for this task is severely limited in the operational environment. Training area geography, route availability, seasonal and lighting effects, cockpit constraints, aerodynamic and safety concerns, and task complexity severely limit the effectiveness of instruction in an operational environment. Given the complexity of the task and difficulties with the operational training environment, methods for designing effective instruction are critical.

c. *Navigation Performance Assessment: Sensory, Perceptual, Cognitive and Overt Performance Indicators*

One of the problems highlighted in the introductory paragraphs is that it can be extremely difficult to assess navigation performance. This section proposes a method for assessing navigation performance that will support a model of expected scan patterns and highlights potential differences between novices and experts. In some navigation tasks, such as commercial flights under instrument meteorological conditions, deviation from the intended route is an adequate measure.

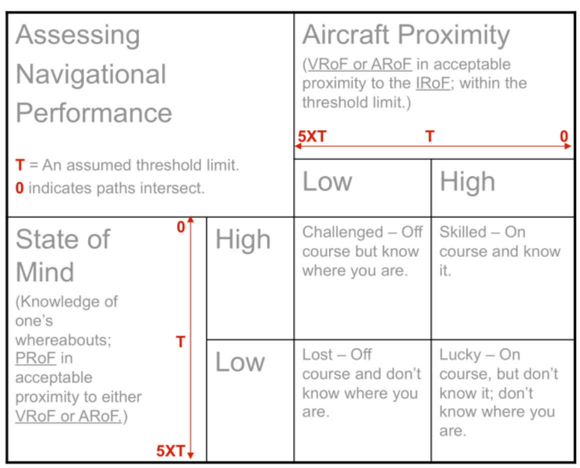

Figure 9. Original Matrix for Assessing Navigational Performance

In helicopter navigation, proximity to a planned flight path tells an incomplete story. Based on terrain, weather and threats, it is often advantageous to

deviate from a straight-line path between check points. Thus, it is also important to consider the individual's state of mind. An individual that is intentionally off the intended route may actually be navigating more effectively than an individual who is unintentionally on course. Figure 9 has been used in previous studies (Kulakowski, 2004; Lennerton, 2004) to capture this difference. In Kulakowski (2004) Lennerton (2004) 'Aircraft Proximity' refers to how close the aircraft is to the intended route of flight. If it is low, the aircraft is not close to the intended route. If it is high, the aircraft is within navigation error standards. Kulakowski, (2004) Lennerton, (2004) use 'State of Mind' refer to the accuracy of an individual's navigation solution. If the state of mind is 'high' the perceived route of flight (PRoF) is extremely close to the aircraft's actual position (ARoF.) The individual knows where they are. A 'low' state of mind refers to conditions where the individual has incorrectly fixed their position.

This work proposes a slight revision to this matrix for assessing navigation performance. Since proximity to the originally planned route does not necessarily correlate to performance and is not likely to drive a difference in strategy or behavior it can be replaced. The revised model considers State of Mind, or navigation accuracy, compared to Confidence. It is depicted in Figure 10.

Assessing Navigation Performance		**Confidence**	
		Low	High
Correctness Perceived and Actual positions match	Low	Struggling. No accurate fix, aware that aircraft is off track.	Dangerous. Lost and doesn't realize it. Positively misidentified correlating features.
	High	On course and lucky. Accurate fix, but not confident in navigation solution.	Skilled performer. On track and certain.

Figure 10. Revised Matrix for Assessing Navigational Skills

The terminology associated with ‘State of Mind’ has been changed to ‘Correctness’. As with the original matrix this corresponds to the degree to which the individual’s perception of the aircraft’s position matches the aircrafts’ actual position. The degree of correctness is compared to the individual’s confidence in their navigation solution. That is, how sure are they of their position? A navigator with ‘high’ correctness and high confidence knows has correctly fixed the aircraft position and is confident of their solution. If the correctness is high but the confidence is low, the navigator could be considered on course but lucky. The aircraft is in the right position, the navigator knows it, but is somewhat uncertain. There may be some ambiguity in the navigation solution. If the degree of correlation between the perceived and actual position is low but the individual’s confidence is high then it is likely they have misidentified some feature. They are confident, but incorrect. If the individual is uncertain of their navigation solution and there is low correlation between where they think they might be and the actual aircraft position, the individual is off course but aware.

Assessing Navigation Performance		Observed Behavior	
		Repair	Maintenance
Optimal Behavior	Repair	Lost, aware and looking. Working to find match. (Once here, novices will stay in this state longer.	Lost and doesn’t realize so is continuing in maintenance mode. (Rare for experts.)
	Mainte nance	Searching for a cue to verify position. (novices will spend more time here.)	On track and maintaining. (Experts will spend more time here.)

Figure 11. Observed versus Optimal Navigational Behaviors

In addition to providing a more meaningful assessment of navigation performance, this matrix also is consistent with a proposed model of scan behaviors associated with these states. This is depicted in Figure 11. In this matrix the relationship of the degree of correctness to confidence maps onto the relationship between optimal and observed navigation behaviors. With poor navigation accuracy and low confidence we would expect to see repair behavior, which would be appropriate based on the

operator and aircraft state. The navigator has no clear idea of where they are but is not laboring under any false assumptions about their position. In the repair mode, individuals would search widely for major prominent features. If the correctness was low and confidence is high, it would be reasonable to expect to observe maintenance procedures when repair would be optimal. Because the individual believes they know where they are, they will likely continue to apply maintenance procedures. If the correctness is high but the confidence is low it would be reasonable to expect the individual to revert to repair despite the fact that maintenance would be more appropriate. In this case, there may be some ambiguity in the navigation solution. Rather than rely on single anchoring features, the individual may scan more broadly seeking a wider set of cues in the out-the-window view and on the map. Finally, if the individual's navigation solution and confidence are high, observed and optimal behavior should both indicate the individual is relying on maintenance functions. The matrix describing novice/expert differences and stages of navigation provides a framework for a more complete model presented in subsequence sections.

d. Developing a Model of Scan Patterns

The basic task of terrain association involves scanning the out-the-window view for a set of one or more recognizable features. These recognizable features could include the unique arrangement of several otherwise nondescript features. To make comparisons with the possible map representations, the individual would encode the out-the-window view; creating an internal representation of what that feature might look like as represented on a contour map. The pilot would then scan the map, executing a pattern-matching strategy. If one or more near matches are found, the pilot would then scan outside the cockpit to confirm their assumption. If the individual was extremely confident in their solution, they may not confirm; they may scan a new section of terrain looking for a new set of distinct features.

This brief description of the helicopter overland navigation task highlights two distinct phases or strategies: repair (or naïve search) and maintenance (or directed search, confirmation.) Individuals reasonably certain of at least one positive match between the current view and a feature represented on the map are likely to execute

maintenance behaviors. Individuals that are uncertain of their location or have little confidence are likely to execute repair or search tactics. Repair and maintenance are likely to occur on both the terrain and map scan.

Both repair and maintenance search strategies will be more efficient for experts compared to novices. It also seems likely that experts will execute the appropriate strategy more often than novices. The probability of choosing the appropriate strategy is likely to be affected by novices' over or under confidence in their ability to find and make effective use of matches.

Expected Performance Based on Navigation Mode and Expertise Level	Novice	Expert
Maintenance Mode (Confirm)	Novices will have increased dwell time, taking longer to capture and encode features in the out-the-window view.	Experts will scan more efficiently, taking less time looking at terrain or map to select salient features.
	Map dwell times will last longer. Pattern matching will take considerable effort. This time/effort works against them as they have less durable intermediate representations.	The ability to more rapidly encode information, store intermediate representations more reliably and for longer times, and make matches will enable efficient shifting of gaze.
Repair Mode (Search)	Experts will efficiently shift between a consistently and appropriately sized set (2-4) of salient features across a wide area of terrain.	Novices will take longer to select and focus on salient features in the terrain. They will select a more variable and less appropriate number of features for comparison. They will tend to focus on smaller regions for selecting candidate features.
	Experts will efficiently overlap scans during successive samples of terrain features.	Novices will not 'anchor' their scan based on most-recent scan. They will inadvertently let known features fall out of their field of view.

Strategy Selection	If overconfident in their solution, novices will rely on single, possibly incorrect matches leading them to execute higher percentage of time in maintenance mode when repair would be appropriate. Conversely if they are unsure of their navigation solution they may execute a higher percentage of time in repair mode looking for candidate matches while overlooking the value of low-confidence, high-accuracy matches.	Higher order skills such as selecting features, encoding them rapidly, preserving them well enable efficient pattern matching. Better ability to assess correctness of their solution will contribute to improved strategy selection. Overall, experts are likely to spend a greater percentage of time in maintenance versus repair when compared to novices.

Table 2. Overview of Novice and Experience Pilot Expected Eye Tracking Characteristics

2. Previous Work Related to Scan and Expertise

Evaluation of scan patterns dates to the 1940s (Duchowski, 2002). Over the last 70 years the underlying eye scan technology has evolved considerably while the range of application areas and motivations for examining scan have expanded significantly. Seminal work provides a basic motivation for using scan to understand expertise and skill acquisition. Recent advances in eye scan evaluation systems, continued evaluation of scan and developing theories of learning and expertise suggest that evaluation of scan patterns in the domain of helicopter overland navigation could provide valuable insight into workload (Di Nocera, Camilli, & Terenzi, 2006, , 2007) and expertise levels (Huemer et al., 2005; Hyönä, Radach, & Deubel, 2003; Juno, Stephen, April, & Robert, 2010; Kasarskis, Stehwien, Hickox, Aretz, & Wickens, 2001; Ottati, Hickox, & Richter, 1999; Reingold, Charness, Pomplun, & Stampe, 2001) (Law, Atkins, Kirkpatrick, & Lomax, 2004; Tien, Atkins, Zheng, & Swindells). Such insight could be used as a guide for development, application and evaluation of instructional systems and could provide an important link to our understanding of learning and skill acquisition.

Bellenkes: Visual Scanning and Pilot Expertise: The Role of Attentional Flexibility and Mental Model Development

In Bellenkes, Wickens et al. (1997) and Wickens, Bellenkes, et al. (1995) measured pilots scan during a 7-segment instrument flight rules (IFR) event conducted in a PC-based flight simulator. The simulator represented a light civilian aircraft and consisted of a set of flight controls and a monitor displaying the instrument panel. There was no depiction of an out-the-window view. This study manipulated the task and protocol to impose varying workload in order to examine differences in instrument scan patterns between novices and experts. Bellenkes, Wickens et al. (1997) concluded that scan data and flight control inputs accurately reflected pilot's ability to allocate resources and achieve improved performance. Based on changes in instrument fixation over different maneuvers and with varied performance criteria, they concluded that experts could extract useful information more quickly than novices. They also concluded that experts relied on a more efficient mental model of the interconnections between and among flight control inputs, aircraft response and flight dynamics criteria. This improved mental model allowed them to allocate visual and attentional resources more efficiently and confidently as demonstrated by scan pattern and flight control inputs.

This work established that eye movement information can distinguish critical differences between novice and expert pilots. Bellenkes et al. (1997) also suggest that these differences can be used to create targeted training programs aimed at developing expert strategies. They do not, however, suggest or recommend how the scan information could be used as input to an interactive real time training system. Additionally the PC-based IFR task is a comparatively closed domain. During instrument flight, the interaction between control inputs, performance indicators and expected outcomes follows a structured mapping. In contrast the domain of overland navigation is much more open. Navigation planning decisions will dramatically affect the difficulty of en route navigation. The interactions of route selection and allocation of attention is much more dynamic and harder to predict. The work proposed here will advance previous studies by examining scan patterns in a more dynamic, open task where interdependencies follow a much more complex and ill-defined mapping.

Kasarskis: 'Comparison of Expert and Novice Scan Patterns During VFR Flight'

In 2001, (Kasarskis, Stehwien, Hickox, Aretz, & Wickens, 2001) extended the work of (Bellenkes, Wickens, & Kramer, 1997) by comparing novices and expert pilots' scan patterns while conducting visual flight rules. They used a PC-based simulation of a basic light, civilian fixed-wing aircraft and compared pilot's scan patterns during landings. Kasarski's work was an important extension of previous efforts as it was the first attempt to compare scan that included an out-the-window view. As with previous studies, Kasarkis found that experts had more efficient scan patterns. Expert versus novice differences applied to both the dwell time required to extract information from the instruments and the sequence of visiting gauges. It also affected allocation of attention between the out-the-window view and cockpit gauges. Examining spatial distribution of scan in the out-the-window view also provided useful insight into novice versus expert differences. Across multiple landings, novices tended to scan a broader and more diverse region of the landing area. Experts scanned fewer points within the landing area. Their dwell points tended to be concentrated on specific critical areas within the landing zone. By examining a VFR task in a consistent area (i.e., the landing area), Kasarkis et al. were able to extend previous work to include out-the-window VFR scan. They were able to make meaningful comparisons of novices and experts in natural scenery. A key factor in making these comparisons and conclusions is that the scenery was constant. The goal of this research project is to extend the analysis to environments where the scenery is variable and different across subjects. During terrain navigation, the scenery will vary based on each navigation decision. Small differences in position along a helicopter overland navigation route can significantly change the set of cues that will be masked or visible. Similarly, allocation of attention creates more variability. Diverting attention from the out-the-window view to the map during a critical phase means that a critical landmark cue may pass by without coming into view of the pilot .

Marshall: 'Identifying Cognitive State from Eye Metrics'

In 2007, (Marshall, 2007) developed a patented algorithm for identifying cognitive state based on eye metrics. In three experiments the Index of Cognitive Activity (ICA) successfully indicated participant's cognitive state, with the capability to

perform analysis in real time. The ICA uses measures of pupil size, eye movements and blinks. It applies two statistical models: linear discriminant function analysis and non-linear neural network analysis. Marshall's work includes describes three key studies. In the first study, participant's cognitive state of relaxed or engaged was evaluated and correctly assessed while participant's conducted problem solving exercises. The second study examined driving tasks. The ICA correctly assessed when participants were focused versus when they were distracted. The third task involved a visual search task. In this last study the ICA accurately indicated differences between alert and fatigued participants.

Together, (Marshall, 2007) experiments and development of ICA provide encouragement for further examination of additional problem domains and varying types of useful information that can be derived from eye tracking data. In particular, the work proposed here will examine eye scan data that can inform an instructional system in the domain of helicopter overland navigation. Marshall's work has proven reliable for indicating relaxed versus engaged, focused versus distracted and alert versus fatigued. Although Marshalls' work is extremely promising for a wide variety of applications, it does not provide insight into the aspects of cognitive state that are useful for cueing instruction. Our work assumes that for engaged, focused and alert participants we will be able to derive behavior-related information that is useful for cueing an instructional system.

3. Applying Eye Tracking to Helicopter Overland Navigation

The idea to investigate the value of eye tracking for uncovering instructional moments was based on observations from previous studies. Since future work should not rely on completing multiple theses, we also investigated other ways to verify the set of neuromarkers that might cue instruction. We found support for the use of eye scan based on the literature reviewed in the previous section and by the cognitive task in Appendix A. In this section, we discuss using the cognitive task analysis in Appendix A to identify neuromarkers that can provide information on trainee's cognitive processes and that would have useful diagnostic value within an instructional system. We consider into internal metal processes such as perception, attention management, level of certainty and workload, which would be useful for instruction.

If we interviewed instructors involved in teaching continuous complex cognitive tasks, it would be easy to imagine they would provide a long list of features of a trainee's internal mental processes they would like to able to measure. Unfortunately, such a procedure has not been formalized. Thus, we opted to use the task analysis in the context of previously reviewed studies on neurophysiological markers (discussed in Chapter II.D.2) to identify elements of cognition that may provide helpful insight.

Ambiguous-match-method	
Analyze-terrain-for-correlating-feature	;a possibly ambiguous feature that because of it's spatial relationship with other features may be used to definitively locate aircraft position
SELECT: possible-correlating-feature-in-view-method	
Estimate-map-representation-of-correlating-feature	
Compare-estimated-representation-of-feature-with-map	
Compare-map-with-feature-to-verify	
SELECT: positive-match-of-correlating-feature-method	

Table 3. A selected portion of helicopter CTA involving choosing and evaluating a candidate feature for navigating when unsure of current position

Table 3 contains several steps in the portion of the CTA associated with resolving ambiguity of a navigation solution. The high-level goal in this stage is to scan the terrain within view for a salient correlating feature. A feature can be considered salient based on any aspect that makes it uniquely identifiable. Features can be uniquely identified based on size, shape, orientation or by position and orientation relative to other features. As an instructor, it would be extremely helpful to know if the trainee was scanning a reasonable set of cues in a reasonable amount of time. It would also be extremely helpful to know if the student felt confident or unsure and if they were overwhelmed or coping well. These

internal factors associated with learning terrain navigation: scan strategy, confidence and workload; suggest a number of potentially valuable neurophysiologic pathways into discerning trainee's cognitive processes. Cortisole levels, heart rate and galvanic skin response might be useful for estimating workload and stress (Schmorrow, Estabrooke, Grootjen, Coyne et al., 2009). Electroencephalogram (EEG) might be helpful for indicating the degree of confidence (Schmorrow, Reeves, Bolton, Campbell, & Schmorrow, 2007). Based on our understanding of expertise and the literature reviewed earlier (Chapter II.D.1.d), we would expect experts to resolve ambiguities much more efficiently and much more quickly than less experienced pilots. The speed and efficiency would also involve a much different scan pattern. Thus, scan pattern analysis should be a reliable indicator of a trainee's speed, efficiency and expertise. An instructor or instructional system that could sense the difference between struggling and proficient individuals would be in a much better position to select and apply timely and appropriate instructional interventions.

Of these potential measures to provide insight into a trainee's internal processes associated with learning, eye scan is likely to provide the most salient diagnostic information and thus drive the most relevant instructional intervention. Knowing that an individual is stressed an instructor could correctly adjust factors that contribute to stress; however, these factors may or may not be helpful for improving training procedures. Knowing that a student is confident is helpful only if you know the student is also correct. Since the only other visible indication of performance—aircraft position—is also ambiguous, confidence is certainly good to know, but doesn't provide diagnostic information. On the other hand, scan information could provide helpful information on two levels. Raw statistics may indicate overall scan efficiency (Pool & Bell, 2004) and thus the trainee's level of expertise. Scan pattern analysis may indicate strategy selection and execution efficiency. The ability to gauge level of expertise and to evaluate a trainee's strategy using eye scan may provide diagnostic value to the instructor.

Since the navigation task is critically dependent on visual tasks, there are numerous elements where scan analysis can provide insight into cognitive state. Table 4 lists one highly visual element that could provide important insight into strategy selection.

GOAL: Scan-for-next-navigation-point	; see cue inventory
SELECT: Follow-hand-rail-method	; usually a linear terrain feature
Positively-identify-hand-rail-feature	
Direct-PAC-to-follow-hand-rail	
GOAL: Update-on-track-progress	
Select-on-track-landmark	
Evaluate-track-deviation	
Visible-intermediate-navigation-point-method	; only if you see the point

Table 4. A portion of helicopter CTA involving selection of a navigation strategy

For the task section listed here, eye scan analysis would be the most likely neuromarker to indicate planning and strategy selection. In using a handrail feature, we would expect an expert to scan well ahead of the aircraft and make frequent visual reference to this terrain element. The expert's scan would likely be well organized using the handrail as an anchor feature to maintain continuity between successive looks out the window. By contrast, novices may not take advantage of handrail features. They would demonstrate a less organized scan, spending more time on extraneous cues and less time relying on available handrail features. Reviewing the cognitive task analysis in the context of available neuromarkers, models of expertise and indications of trainee's internal processes that could provide useful diagnostic feedback, indicates that scan pattern may be a useful neurophysiological marker for cuing instruction.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODS

A. RESEARCH QUESTION

Complex cognitive tasks present a unique challenge for those tasked with designing and applying training solutions. While simulation technology continues to provide increasingly realistic representation of real environments, high fidelity recreations are not likely to reach full potential effective training platforms unless they can respond on the fly. Effective training for complex tasks requires more than a recreation of the task in a simulation environment. It requires an effective design process; preferably one that does not require validation via extensive training effectiveness evaluation.

Previous studies (Ross, Phillips, Klein, & Cohn, 2005) connecting cognitive models of expertise with training provide promise for tailoring simulation to groups of individuals based on their level of expertise. However, there currently is very little research connecting methods for assessing expertise in real time with instructional design for simulation. For the representative task of helicopter overland navigation, will real time analysis of scan pattern provide sufficient insight into a trainee's level of expertise to adequately inform an instructional system?

The goal of this research is investigate neurophysiologic cues into behavior and cognition as they relate to the development of expertise on a complex cognitive task. The aim is to determine if analysis of eye scan data collected during a terrain association navigation task can provide sufficient insight into a novice trainee's behavior and cognition to cue an instructional system or indicate requirement for additional neurophysiologic markers. The first part of this work will compare statistical characteristics of scan. Parameters examined include allocation of scan time between out-the-window and map views, mean dwell time and dwell duration frequency distribution. Previous work suggests that experts will divide their scan time more efficiently; fixating on fewer, more salient features with overall reduced mean dwell time and higher frequency of shorter dwells. The remainder of this work involves qualitative analysis of the temporal aspects of scan data. The goal of this analysis is to provide

additional support for the application of real-time neurophysiologic cues to improved instructional systems. The first step in this phase involves identifying unique characteristics of expert's eye movement that lead to success. Novice habits will then be compared to experts to identify opportunities to cue instruction.

B. EXPERIMENTAL HYPOTHESIS AND DESIGN RATIONALE

This section provides background on methods selected to validate the hypothesis, including rationale behind selection of eye scan metrics and expected differences between novices and experts. The overall goal was to create a simulation environment where the key cues from the real world were recreated in sufficient detail to allow participants to follow real-world protocols and procedures in task execution. Designing the experiment involved striking a balance between the task difficulty and skill level of available participants. Steps to adjust task difficulty and simulation design decision are discussed in detail

The high-level research goal is to determine if statistical properties of scan will distinguish novice and experts in continuous complex cognitive task of navigation via terrain association; and if scan visualization will indicate underlying strategy of navigators. The experimental hypothesis is summarized in Table 5 .

Dependent Measure	Hypothesis
Performance Measures	
Flight path RMS error	RMS is a poor indicator of expertise.
Difference between actual and Ideal flight Time	Experts will come closer to arriving on time.
Basic Dwell Characteristics	
Dwell duration – median	Experts' median dwell will be less than novices.
OTW dwell – median	Experts' median OTW dwell will be less than novices.
MAP dwell – median	Experts' median map dwell will be less than novices.
Dwell duration – mean	Experts' mean dwell will be less than novices.
OTW dwell – mean	Experts' mean OTW dwell will be less than novices.
MAP dwell – mean	Experts' mean map dwell will be less than novices.
Dwell duration – STD	STD of expert dwell duration will be lower than novices.

OTW dwell – STD	STD of expert OTW dwell duration will be lower than novices.
MAP dwell – STD	STD of expert map dwell duration will be lower than novices.
Higher-level Scan Characteristics	
Percentage of flight time in Dwell	Experts will have more saccade time, thus shorter overall dwell time.
Fixations per OTW view	In each OTW view, experts will fixate on fewer points
Fixations per MAP view	In each map view, experts will fixate on fewer points
View Changes/Flight Time	Experts will change views more frequently than novices
OTW Dwell / Total Dwell	Experts will spend more of their dwell on OTW than map

Table 5. Summary of Experimental Hypothesis

1. Task Difficulty and Target Group Selection Interaction

The goal in recruiting subjects and tailoring the experimental conditions and navigation task is to demonstrate a distinct difference in performance between two groups categorized as novices and experts. Neither the model of expertise (for example Dryfus' Staged Model of Expertise) nor the specific categorization of an individual within that model (novice versus competent beginner) is useful for the purposes of this study. The primary goal is to explore the relationship of traditional metrics of navigation performance, such as difference between intended and actual paths, and the utility neurophysiologic markers. We anticipate that comparing performance metrics and selected markers may provide insight into cognitive processes associated with the acquisition of expertise. Ideally, the subject pool would represent two very distinct stages of performance. Figure 12 depicts the ideal distribution of subjects based on navigation performance. Given a large subject pool, one way to achieve an ideal distribution of subjects would be to select the bottom one third and top of the subject pool based on likely indicators such as overland flight experience and flight qualifications. For follow on studies, training and operational squadrons would provide a much greater opportunity—in terms of the number of subjects available, likelihood of seeing a variety of strategies, and for assessing potential instructional value. However, since the scope of this study is limited to an investigation to indentify neuromarkers and their potential

utility for cuing instruction, the controlled laboratory environment is preferable. Limitations on recruiting a large subject pool are likely to make selection of only the top and bottom third of the available subject pool impractical. To compensate for limitations in available subject pool, we varied the task difficulty to improve the likelihood that we would find a discernable difference in performance of our subject groups.

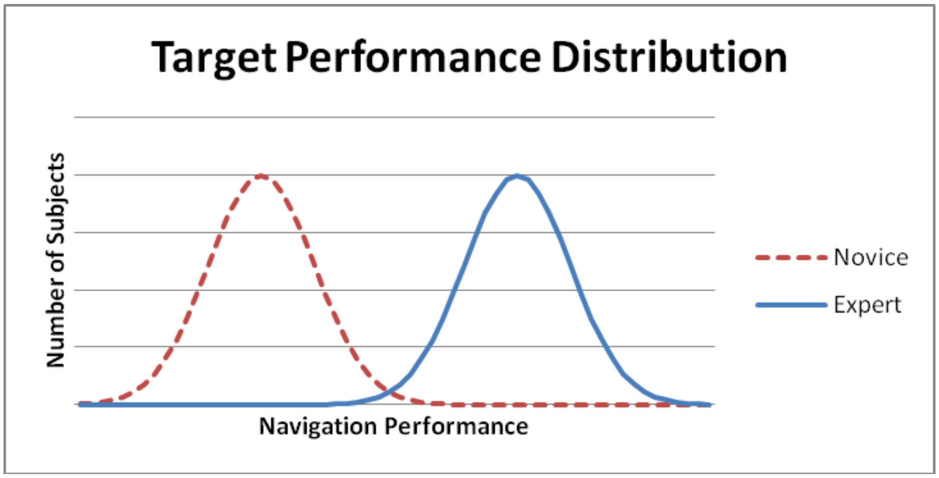

Figure 12. Desired Performance Distribution of Novices and Experts

There are a number of options for adjusting the difficulty of the task to achieve desired distribution of performance. Options considered included: navigation route, visibility, altitude, groundspeed, and addition of concurrent tasking (including difficulty of aircraft control.)

Terrain and route selection have the greatest impact on task difficulty. Flat, featureless areas provide few distinct navigation cues. Although such regions may provide a difference between novices and experts, they will provide very limited ability to vary the difficulty of different segments of the route. On the other end of the spectrum, rugged terrain provides numerous cues and increases the opportunities to vary the difficulty of the route. Section of an area with rugged terrain improves the opportunities to select route segments where expert performance will be clearly delineated from novice performance. The area for selecting routes will involve rugged terrain to facilitate selection of segments that will create clear difference in novice and expert performance. More difficult segments will be selected with salient, readily

distinguished features in view for a relatively short period. Easier routes will be selected with distinct, prominent features in view for longer periods.

Navigation is clearly more difficult in restricted visibility. As a secondary method to vary navigation task difficulty, a means for adjusting the terrain visibility will be provided. This will enable a more fine-grained control of task difficulty. The method for controlling the aircraft during the navigation task also impacts task difficulty. In the aircraft, the non-flying pilot is responsible for providing verbal direction to the flying pilot to maintain position along the intended route. Ideally this protocol could be duplicated in the experiment. Unfortunately this would introduce variability that would be too difficult to control. For example, verbal directions from the non-flying pilot may make reference to terrain features. The ability to provide a clear verbal description of a unique feature is a learned task that may be separate from the navigation skills we are interested in studying. By enabling verbal control, we could possibly end up measuring communication skills.

2. Terrain Model and Navigation Route

As described above, selection of terrain and route generally have a significant impact on the task difficulty. The navigation route was specifically tailored based on the expected range of participants. Selecting the route for the participants rather than allowing them to plan their own route is somewhat artificial. As previously noted (Chapter II.D.1), navigation is never the exclusive objective of the mission. While specific points related to the mission, such as medical evacuation landings zones, are often assigned; crews are normally responsible for selecting the most appropriate connection of intermediate waypoints based on threats, terrain, anticipated weather and other factors. Since route selection can impact route difficulty, there is at least some chance that individuals may have elected to navigate to the objective using a different set of key landmarks and routes based on personal preference. To minimize the chance of this, we conferred with multiple subject matter experts in terrain navigation while designing the route. Several candidate routes were created and evaluated informally in the laboratory setting.

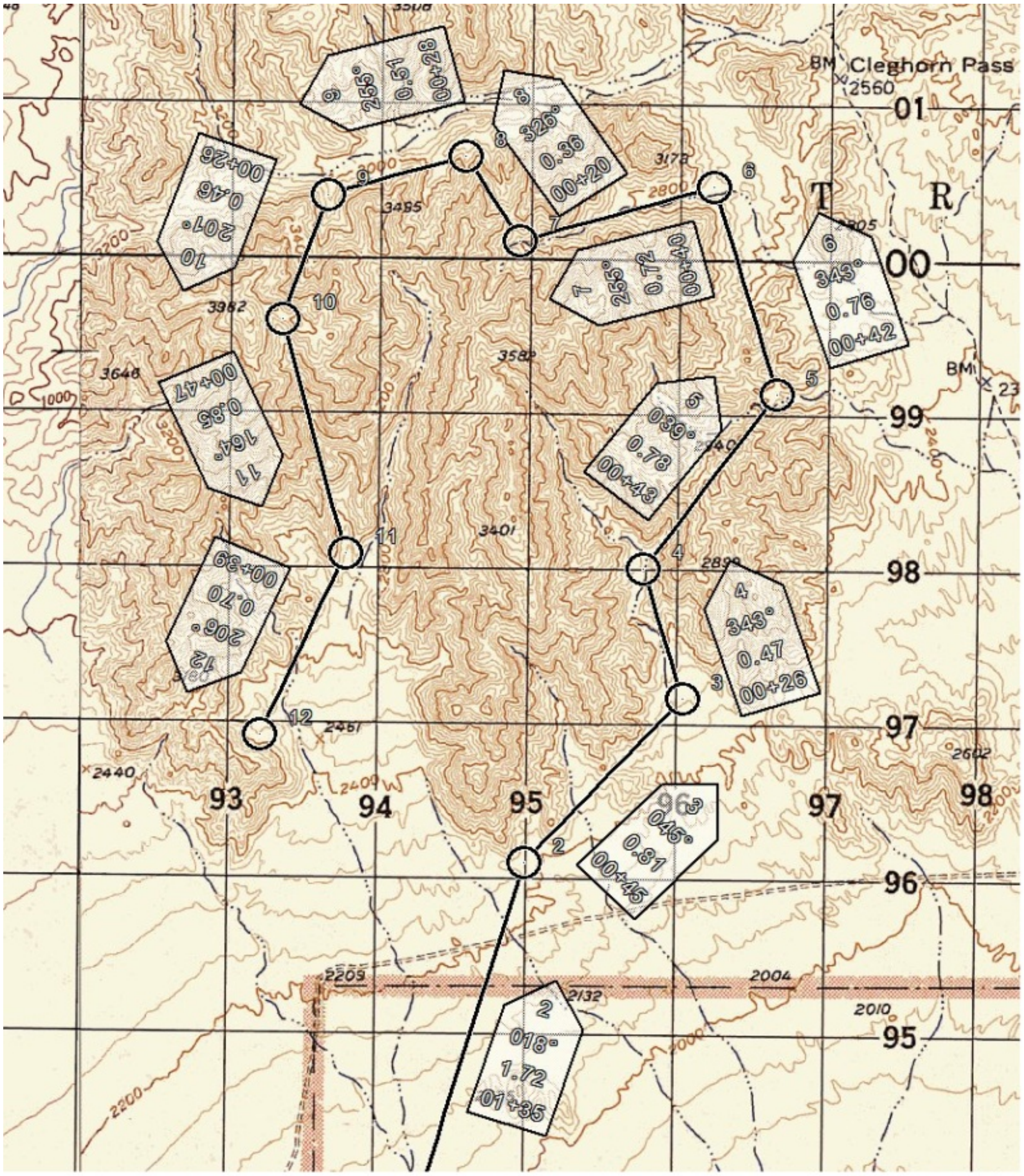

Figure 13. Experiment navigation route

The route was designed to be moderately difficult for the experience level represented by the student population at the Naval Postgraduate School. Students at NPS normally have completed several operational tours. They could be expected to perform well on fairly challenging routes. The 29 Palms area provided rugged terrain similar to operational environments and devoid of vegetation. The route consists of 11 legs, each about kilometer or at planned speeds 40 seconds long. The route was designed with

suitable catching features to minimize the chance of participants wandering significantly off course. Waypoints were selected with varying levels of difficulty

3. Out the Window Display

Part of the experimental design was based on a tradeoff between eye tracking equipment calibration complexity and representational fidelity of the simulation. Based on previous studies (Hahn, 2005; Kulakowski, 2004), we know that navigation in a terrain virtual environment correlates to terrain navigation in a real environment. Ideally the study would have recreated all of the characteristics of the real environment. Recreating the complete field of view available in the aircraft and calibrating eye tracking equipment was not feasible. Thus, we decided to limit field of view to a single display. The limited field of view in the simulation as compared to the aircraft should impact novices and experts similarly. The focus of the study is not a comparison of how experts and novices perform in the real world. It is closer to how do experts and novices compare in their use of available field of view to navigate.

4. Helicopter Motion Control

As detailed in Chapter II.E.1, during a typical mission the non-flying pilot is responsible for navigation-related tasks. The previous section described the difficulties associated with providing a verbal interface for aircraft control. To avoid potential task overload and distraction associated with using verbal control methods, the research system includes a joystick operated by the navigator for flight control. The goal in selecting a flight control and dynamics model was to provide a system that was very easy to learn and use and that would generally conform to the flying characteristics of an operational helicopter. Acceleration rates, velocities and turn rates should all reasonably approximate a helicopter without modeling all the complexities of actual flight. High cognitive load associated with tasks other than navigation could result in navigation task performance degradation and difference in scan patterns not related to navigation. It is likely that a flight control system with high cognitive workload would have a greater effect on novices compared to experts.

A secondary goal was to provide a vehicle motion model that would be easy to recreate in other laboratory settings and for other simulation configurations. Potential future work includes migrating to a more challenging version of the task, closer to full task complexity. Other potential future work includes the possibility of conducting evaluations in a higher fidelity simulation platform. Providing a scalable flight dynamic model would minimize differences across this and future research and improve the chances of making meaningful comparisons across these studies.

Previous studies (Kulakowski, 2004) made use of a key-board based approximate flight model. The height above terrain and groundspeed were fixed. The subjects used a modified set of verbal commands to initiate either a half-standard, or full standard rate turn and return to level flight. While the verbal control and keyboard model was effective, there were several shortcomings. It could not easily be replicated in other laboratory settings. Additionally it did not scale to support simulation at varying levels of difficulty. Most critically, the physically-based model limitations on turn rate (half standard or full standard rate) and turn acceleration rate combined with the verbal control metaphor reduced the operator's degree of fine-grained control and navigation accuracy.

To support improved fine-grained control, increase portability and scalability while retaining reasonable aerodynamic properties and ease of learning and use, the research system used in this study was a modified version of the commercially available X-Plane. To reduce cognitive load associated with flight dynamics, the modified X-Plane system incorporates automated terrain following and airspeed hold features. The operator is provided joystick that directs roll. Based on joystick input and flight dynamics model the X-Plane controller determines aircraft position and orientation. The newly calculated position is then artificially adjusted to maintain constant groundspeed and height above terrain. This design retains important flight characteristics including pitch, roll and yaw response rates and turn radii.

There were no specific tests of either the degree of realism or the degree of cognitive load imposed by the flight dynamic model. Informal usability studies by developers and lab personnel, including several with significant rotary wing aviation experience, provided confidence that the system did not impose significant cognitive load

and that it represented reasonable flight dynamic properties. Given that the user only controlled rotation about longitudinal axis, roll response and turn radius were viewed as critical. In particular we were concerned that subjects would be able to predict turn radius so that turns could be initiated in time to conform to intended flight path. Although there were no specific tests for difficulty associated with controlling the flight model, subjects were provided with an open-ended practice session to become familiar with the flight dynamics. During the introductory section, individuals were asked to follow a navigation route depicted in the out-the-window view. They were also directed to turn to specific landmarks similar to those they would find along the navigation route. Since users were provided an open-ended practice session, it is assumed that individual's self-reported comfort level would demonstrate suitability of the flight dynamic system. If users can reasonably control the flight model to conform to a desired flight path and require no more than 20 minutes to become comfortable, the model fills the intended role.

5. Cockpit Displays

In addition to the map display, there are several other cockpit gauges used in the navigation task. To support dead reckoning navigation, the simulation needs to present timing, heading or ground track, and ground speed information. These data are normally provided by a conventional analog clock with resettable sweep hands for elapsed minutes and seconds, a compass—normally integrated into a horizontal situation indicator (HSI) and an airspeed indicator. Often groundspeed is provided by a digital readout on the HSI. To support navigation by altitude correlation, the simulation would also need to provide a barometric altimeter.

The goal in selecting and configuring the simulation cockpit displays was to provide the required information so that it was readily accessible and did not require significant time to learn the location of specific gauges. Providing a full set of gauges would more closely resemble the aircraft task, however this would introduce the possibility that performance differences would be based on the fact that some subjects may be better at adapting to, or already familiar with a given configuration. Since novices tend to take longer than experts to extract useful information from display gauges, this would likely have a stronger negative effect on novices than experts. Thus

the simulation used in this study only contains gauges that are specifically useful for navigation: a clock with resettable minute/second sweep hands, a compass, airspeed indicator and barometric altimeter. Since groundspeed displays tend to be very specific to individual type, model, and series cockpits, the instrument cluster includes a conventional airspeed indicator. Aircraft—specific ground speed readouts could give a disproportionate advantage to subjects familiar with that configuration. Developing novel groundspeed readout would potentially increase the time required to learn the system. Since groundspeed is automatically controlled and subjects are briefed that groundspeed is clamped providing indicated airspeed versus groundspeed information should not affect navigation ability. The potential for negative training transfer, including training pilots to ignore ground speed information is discussed in Chapter II.D.1.b. Reducing the complexity of the instrument panel also introduces the potential for negative training transfer. Presenting an oversimplification of the task may result in depressed overall learning and poor far transfer (R. Clark, 2003). This investigation is focused on improving simulation to build a better sense of when to provide or remove training simplifications such as automated ground speed control and simplified instrument representations.

6. Map Display and Control

Display and control of map information provided a significant design challenge centered on providing reasonable similarity between the real world and virtual task whilst providing opportunities to collect vital user information. Cockpit map displays vary widely across DoD helicopter communities with very little commonality in the availability of moving map and tactical displays. The single element in common is the conventional paper map. Ideally this project would have followed the standards applied in DoD helicopter navigation training squadrons such as HT-18. For low-level navigation, the map normally used is a 1:50,000K scale topographical land map (TLM).

Unfortunately, providing subjects with a paper map would have made it exceedingly difficult to gather eye movement data during map scanning; in part because the location of the map would not be consistent. Although some pilots keep the map on or near a kneeboard or other constrained area, many pilots opt to hold the map in front of

them to minimize the time to transition from an out-the-window to map scan. Because of the small map scale in relation to speed and constraints on cockpit space, maps are normally folded over several times for one navigation route. Pilots are also trained to rotate the map in the direction of travel further complicating the task of gathering eye movement data related to map scan.

Providing a paper map and constraining how it could be used was considered. Providing a navigation route that conveniently fit on a single, reasonably sized map to conform to cockpit space and fixing this map in place would solve some of the issues. A paper map fixed in place would allow collection of eye movement information on the map. Unfortunately it would not address the issue of map rotation. Without a means to account for map rotation map horizontal and vertical (x and y) intersection point could be calculated, but there would be no way to correlate the latitude and longitude, thus terrain feature, being scanned.

To meet the primary goal of representing the task in a manner consistent with real world differences in novice and expert performance while providing simulation instrumentation opportunities, the final design involved a digital map. The digital map used here replicates the standard CADRG maps used by DoD communities for both planning and operations. While the digital map maintains correlation between real world training and operations and the synthetic environment, it does not resolve the issue of map rotation. With a paper map, orientation to track up is an intuitive and trivial action. Ideally the digital recreation would provide a similarly intuitive and trivial means to rotate the map. Unfortunately, this would have increased task complexity; potentially resulting in creation and evaluation of a task not closely correlated to the real world analog.

To remove this possibility, the digital map integrated an automatic track-up maintenance feature. This solves the issue of providing an instrumented map display in the synthetic environment, however it increases the level of abstraction between the real and synthetic environment, potentially favors subjects with more familiarity with moving map displays and creates the possibility for negative training transfer. Reinforcing the habit of rotating the map in the direction of travel is part of the normal training regimen.

This habit requires that the navigation pilot maintain consistent awareness of maneuvers initiated by the flying pilot. Providing an automated track-up capability for the map removes the requirement to maintain consistent awareness of aircraft heading and reduces the opportunity to practice manually rotating a map as required in some DoD helicopters. Unfortunately there is no clear means to more faithfully represent the task, to include map scrolling and rotation while providing feedback on user map scan. In balance, this project assumes that these differences will impact both novices and experts similarly. There were no specific measures available to determine the degree to which this may impact user's performance or the potential for negative training transfer.

7. Evaluation and Data Collection Considerations

As noted in the introductory paragraph and further detailed earlier in Chapter 0, assessing navigation performance is difficult. Ideally an instructor would base instructional interventions on the relationship between the trainee's degree of correctness and their degree of confidence. Degree of correctness is measurable to a certain extent by observing proximity to intended route. A trainee's degree of confidence is much more difficult to assess. Several options for evaluating confidence were considered; these included verbal protocol, post-navigation task self-assessment or tests and addition of concurrent, navigation-related tasks. Of these methods considered, post navigation task self-assessment and tests were included.

Verbal protocol can be a very positive indicator of both how effectively an individual is navigating and how confident they are of their navigation solution. For assessing accuracy, navigators that can describe their position precisely in terms of nearby terrain features must have an accurate navigation picture. The converse is not always true. A highly skilled pilot with an accurate internal description of their location may not be able to encode this information and provide a verbal description that is clear to an evaluator.

Consistent with cognitive load theory discussed in Chapter 0, it is also important to consider the type of cognitive load this would impose and how this might affect novices and experts differently. The process of verbally encoding information on location and providing this description may detract from the primary task of navigation.

Since in its basic form, navigation does not require verbalization of location, this not intrinsic cognitive load. If it constitutes germane cognitive load it could lead to a training effect and thus could present a confound in the experiment. If, as seems likely, providing verbal protocol would constitute extraneous cognitive load, it would unnecessarily complicate the task. Verbal protocol would likely affect novices more than experts. Novices would be expected to have weaker encoding and verbal description as well as meta-cognitive, task-allocation ability. While having subjects provide a running verbal description of their progress would drive a difference in performance outlined in Chapter II.D.1.d the differences may not be related to basic navigation ability. The differences in performance may be attributable to a more diverse set of higher level skills (encoding and verbal description, meta-cognitive.)

Assigning a concurrent task related to navigation could feasibly be done to assess navigation accuracy and confidence. As with verbal protocols, concurrent tasking would then depend on the individual's ability to integrate this task into the overall workload. Novices, who are generally prone to incorrectly estimate their own ability would also be likely to have more difficulty appropriately allocating mental resources and selecting among secondary tasks. Thus, as with verbal protocols, the study would risk evaluating a larger, aggregate set of tasks rather than measuring the basic fundamental navigation task.

Previous related studies (Schmorrow, Cohn, & Nicholson, 2009) have made effective use of post-trial questionnaires and assessment techniques to provide insight into subject's level of ability and confidence. Although this information is subjective and may not provide an analytic basis for conclusions, it will not interfere with the basic task and will not add to the complexity of the experimental design. Details of the questionnaire and assessment are included in Appendix B.

Camera placement, lens selection and calibration were critical for maximizing eye tracking accuracy while maintaining real world task similarity. Navigation in a helicopter requires an active scan and frequent changes from the out-the-window to the map view. Therefore the system needed to support natural freedom of movement. Eye tracking accuracy is a function of the size and consistency of the tracked features; this is difficult to attain when a user is frequently changing position and orientation.

C. MEASURES

This section describes the independent and dependent measures. Our definition of expertise used as an independent variable is based on the discussion of deliberate practice in Chapter II.B.3.c. The list of dependent measures is developed based on previous eye scan literature reviewed in Chapter II.D.2, the theoretical model developed in Chapter II.D.1.d and the cognitive task analysis discussed in Chapter II.D.1.a.

1. Independent Variables

The independent variable for this study was expertise. Expertise was based on two user characteristics: instruction experience and total flight hours. Based on the discussion of Ericsson's work, we assumed that flight hours alone are insufficient to capture expertise. According to (Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson & Lehmann, 1996), expertise is defined as years of deliberate, focused practice, generally at least 10 yrs. Therefore, flight hours by itself gives no indication as to whether focused and deliberate practice occurred. Similarly, level of instructor probably does not indicate years of experience. Thus, a combination of total flight hours and instructor experience level may provide a more accurate representation of participants' true expertise level.

2. Dependent Variables

Based on the model developed in Chapter II.D.2, Table 6 lists the dependent measures and their definitions.

Variable Name	**Description**	**Formula**
Actual versus Ideal flight time	Difference between the ideal time to fly and the actual time from waypoint two to waypoint five	$t_{wp5} - t_{wp2}$
Flight path RMS error (ft)	Root Mean Square error between the actual flight path and the ideal flight path. Smaller RMS error means the pilot flew closer to the way points. Ideal flight path was	$\sqrt{\dfrac{\sum_{k=t_0}^{t_f}\left(x_a(k)-x_o(k)\right)^2}{n}}$

	defined as a combination of straight line segments between consecutive way points.	
Total scan duration (sec)	Sum of dwells on both OTW and MAP = Total flight duration – total saccade duration. Assuming the pilots cannot extract valid visual information during saccade, total scan duration may tell us how long the effective visual scan was.	$\sum_{k=1}^{n_{total}} dwell(k)$
Percentage of Flight time in dwell	Total scan duration / Total flight duration * 100	
MAP scan duration (sec)	Sum of dwells on MAP	$\sum_{k=1}^{n_{MAP}} dwell(k)$
H/D*100	MAP scan duration / Total scan duration * 100	
Dwell duration - median (msec)	Median value of all dwells. Used `lognfit` in MATLAB to estimate the parameter.	lognfit ({dwell})
OTW dwell - median (msec)	Median value of OTW dwells	lognfit ({OTW dwell})
MAP dwell - median (msec)	Median value of MAP dwells	lognfit ({MAP dwell})
Dwell duration - mean (msec)	Mean value of all dwells	$\frac{\sum_{k=1}^{n_{total}} dwell(k)}{n_{total}}$

OTW dwell - mean (msec)	Mean value of OTW dwells	$$\frac{\sum_{k=1}^{n_{OTW}} dwell(k)}{n_{OTW}}$$
MAP dwell - mean (msec)	Mean value of MAP dwells	$$\frac{\sum_{k=1}^{n_{MAP}} dwell(k)}{n_{MAP}}$$
OTW-MAP view changes	Total number of view changes between OTW and MAP, i.e., every time the pilot moves his gaze from OTW/MAP to MAP/OTW, this metric increase by one.	$\sum diff(view\ index)$
P/B (numbers/sec)	OTW-MAP view changes / Total flight duration	
Dwell duration - STD (msec)	Standard deviation of all dwells	
OTW dwell - STD (msec)	Standard deviation of OTW dwells	
MAP dwell - STD (msec)	Standard deviation of MAP dwells	

Table 6. Dependent measures

D. APPARATUS

This section is divided in two parts. The first subsection contains a description of the hardware and software used to recreate the navigation task and collect participant's scan information. The second subsection provides a description of the software used for qualitative analysis of scan.

1. Helicopter Navigation Simulation

A detailed schematic of the hardware and software is shown in Figure 14 and a picture of the laboratory set up is show in Figure 15. The basic elements of the apparatus

included cockpit-style seat with side mounted joystick, a 110cm by 61cm display to present an out-the-window (OTW) view, a 88.5cm by 50cm display for the map and instrument display, cameras for collecting eye data and associated personal computers for driving the displays and collecting data.

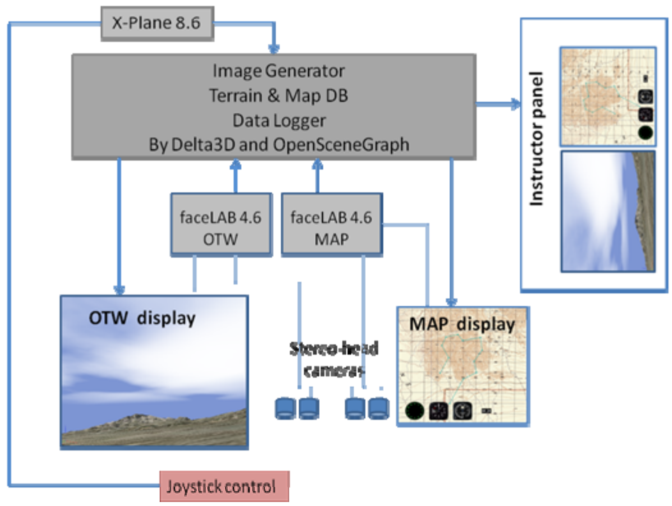

Figure 14. Schematic Diagram of Experiment Setup

The out-the-window monitor was positioned 87 centimeters in front of the participant and covered 65 degrees of the field of view. The rendering software presented 65 degrees field of view. The map and instrument display monitor was positioned 72 cm to the right at a 75 degree angle relative to the user. The map display was positioned as a close approximation of aircraft configuration that would facilitate collection of eye tracking data. The map and instruments were displayed on approximately the leftmost ½ of the monitor. At a distance of 72 cm from the user the map and instrument data occupied 32 degrees of the user's field of view.

Figure 15. Equipment Setup

The user controlled the aircraft position using the joystick. Joystick inputs were used to drive an aerodynamic model that was designed to provide simplified, intuitive control and provided characteristics of realistic flight. The aerodynamic model provided an automated terrain-following feature at a fixed 150 above ground level (AGL.) The aerodynamic model also maintained a fixed 60 knots groundspeed. The user had control of heading only. Heading change was accomplished by laterally displacing the joystick left or right. Left and right displacement was proportional to roll angle. The view was calculated based on accurately representation of ego centric roll. With ego-centric roll representation, the terrain model would appear to roll to the right in a left turn. In a left turn the terrain model would appear to roll right. Forward and aft stick displacement resulted in no change to the aerodynamic model. Fore and aft stick displacement resulted in changing the pitch angle of the viewpoint. Using forward and aft stick displacement to control pitch angle provided users the opportunity to pull back on the joystick when flying into steep terrain to see the top of terrain features. It also provided users the opportunity to pitch the viewpoint down to keep terrain features in view when flying down steep terrain.

The map display presented a 1:50K topographical land map (TLM) typically used for flight planning and execution. The map uses a standard Compressed ARC Digitized Raster Graphics representation. The map was fixed in position about the pair-wise mean

of the waypoints that comprised the familiarization and trial navigation routes. The map orientation was synchronized to the aircraft's heading to maintain a track-up orientation. The map displayed numbered circles to represent the waypoints that made up each navigation route. The bottom portion of the screen contained instruments to support the navigation task. The left-most instrument display was a compass typical of legacy Navy H-60 (SH/HH-60F/H) displays. To the right of the compass display was a typical barometric altimeter and radar altimeter. The rightmost portion of the instrument cluster contained a digital-style elapsed time clock.

The eye data collection system consisted of two FaceLabs stereo camera pairs with associated laptop and software systems. One set of cameras was associated with the out-the-window display and a separate, independent set was associated with the map and instrument panel display monitor. The OTW camera system was centered in front of the user on a telescoping platform. The telescoping platform simplified the process of adjusting the camera position for each user. Any change to the camera position and orientation relative to each other or to the horizon required the camera stereo-head to be recalibrated. Thus, changing the camera's view of the participant via the telescoping platform decreased the complexity and time required to set up for individual participants. The cameras for the OTW view were offset slightly to the right. That is, rather than center users in the camera frame according to the vendor's recommendation; participants were offset slightly to the left side of the camera frame. Offsetting the participant slightly from center of the tracking cameras accommodated for participant's tendency to lean toward the map display on their right.

The camera pair for the map view display was mounted parallel to the map display and offset forward, toward the out-the-window view display. Figure 15 shows the general layout. The depicted camera layout accounted for participant's natural posture and scan habits while seated in the cockpit seat. Participants would naturally scan the map by turning their head to the right. The cameras were positioned forward relative to the map display to provide camera viewing angles nearly perpendicular to the participant's face. The camera postion improved tracking accuracy while allowing unconstrained head movement. More typical usage would call for the camera system to be mounted in the center of the display being tracked with nearly symmetrical

configuration. Although the configuration described here is not considered ideal according to vendor literature, tracking accuracy achieved was very close and occasionally better than the more conventional tracking geometry used for the out-the-window view.

Figure 14 depicts the software architecture that renders the out-the-window and map view based on participant's joystick input, collects and manages eye tracking information and handles data collection. Joystick input is picked up by a small form factor PC that handles aerodynamic model calculations. The aerodynamic-model PC runs X-Plane version 9.0. The software interfaces with X-Plane to provide realistic aerodynamic characteristics of roll rates, radius of turn, and consistent airspeed. The X-Plane aerodynamic model is set to maintain altitude hold at 7,000 feet mean sea level (MSL) to ensure the model does not collide with the terrain model. Pitch information from the joystick is used to control the pitch view angle in X-Plane. Position, orientation and view angle information from X-Plane is broadcast to the image generation personal computer.

The image generation PC adjusts the X-Plane provided position and orientation information to 150 feet above the terrain model. The resulting terrain view is rendered on the out-the-window display. When in practice mode the IG software could optionally render large (10 meter) spheres 150 feet above ground level at navigation waypoints. The spheres were provided to allow participants time to gain familiarity with the joystick control and map rotation without being overly concerned with navigation task. Aircraft flight model heading information is used to rotate the map display to maintain track up orientation. Software on the image generator PC also update and the compass, barometric altimeter, radar altimeter and elapsed time clock.

Each eye tracking system is handled independently. Each camera pair is connected to a single laptop running Facelabs software version 4.6. To maximize facial features visible and to allow for appropriate range of user motion, the 12.5 mm lenses were used. Facelabs system can track eye data based on IR-illuminated facial and eye features (classic mode) or based on the reflection on the pupil of an IR-strobe light synchronized to the frame rate of the camera (precision mode.) Although precision mode

generally offers improved accuracy, previous work (Jungkunk, 2009) and the requirement to use multiple systems drove the requirement to use classic tracking mode. Eye tracking data is recorded at 60Hz. Once calibrated and enabled, the system provides myriad tracking data via network connections. The two systems independently streamed relevant tracking data to the IG PCs. Additionally, all eye tracking data was recorded on the Facelab laptop for subsequent retrieval and analysis.

Within the IG software, eye tracking data from the map and out-the-window view was combined and merged with aircraft position data. Recorded data from the IG system included aircraft position and orientation. The IG software could optionally depict real time scan intersection data for both the map and out-the-window views. For the map view a blue dot was depicted on the map display at the point of eye gaze intersection. For the out-the-window view a yellow sphere was displayed at the point where the eye gaze intersected the terrain model. The IG software also allowed the operator to select from several different navigation routes. The IG software can also be run in after action mode. In after action mode mode a mission could be replayed with the participants gaze intersection data depicted in the terrain scene and on the map.

2. Scan Visualization Software for Qualitative Analysis

The goal driving development of the visualization tool was to provide a representation of spatial and temporal correspondence between features scanned in the out-the-window and map views in relation to the actual and perceived aircraft location. Initial efforts concentrated on a map view which depicted the intended route, the aircraft track and a variable time-based window of recent out-the-window and map scan points. By adjusting the window of time display and updating depicted map and out-the-window scan intersection points on the map, observers could easily detect basic differences in scan patterns. The left side of Figure 16 depicts an expert's efficient scan with close overlap between the map and out-the-window view. The right side depicts a novice's more varied and less coherent scan pattern.

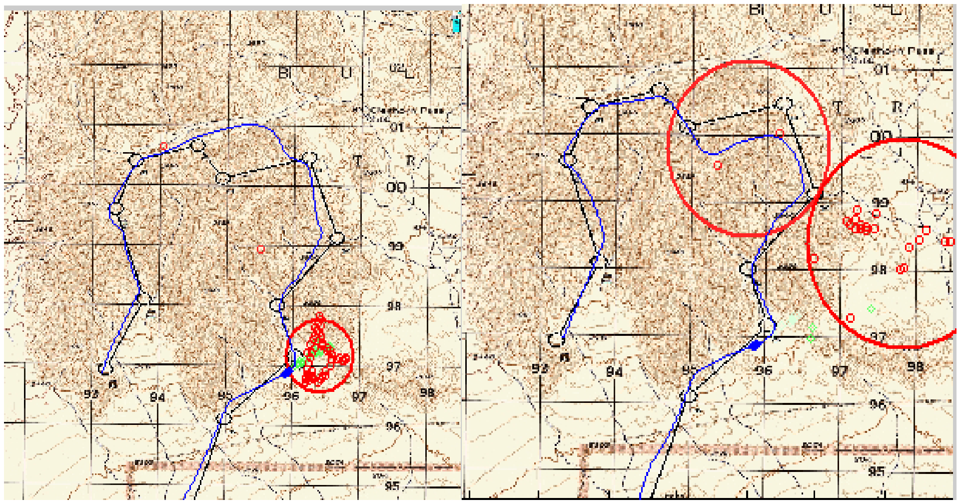

Figure 16. Blue—aircraft position, Green—terrain scan, Red—map scan. On the left an expert terrain and maps scans overlap. On the right a novice's scan is less coherent and overlaps less

While depicting out-the-window and map gaze, intersection points were extremely helpful for exploring differences in correlation between spatial and temporal map and out-the-window eye scan data, it omitted critical information. Without a depiction of the visible terrain and the participant's scan of that terrain, it was difficult to evaluate if critical cues in the out-the-window scene were included or omitted. Therefore subsequent revisions of the tool added the capability to review and replay the participant's experience in an instrumented virtual re-creation of the task. Figure 17 shows an instrumented virtual recreation of a participant's performance.

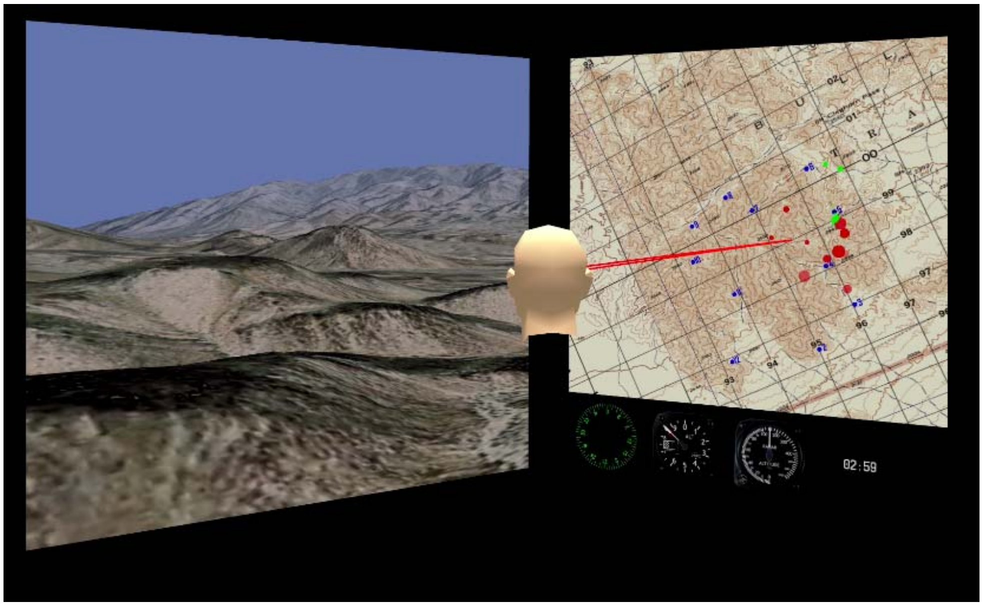

Figure 17. Screen capture of scan visualization tool. The map on the right indicates scan. Green indicates out-the-window dwell location, red indicates map scan dwell points

The instrumented version recreates the operator's interactions to include scan data during system use. The pilot's gaze direction is indicated by two vectors originating at the operator's eyes. Temporal out-the-window and map scan information is presented on the map in the lower right corner of the application. The virtual environment replay allows the operator full control over their view point. The replay provides typical video type controls: rewind, pause, stop and fast forward. All of the variables displayed are available in real time. Network bandwidth and application CPU demand allow for processing to be conducted in real time. In contrast, the current training environment provides very little opportunity for trainers to derive even casual point sample data from a trainee's scan pattern. Evaluation as a real time feedback and evaluation tool remains a promising area for future work.

E. PROCEDURES

Participants were briefed on the overall procedures following a script listed in Appendix B. They were informed that the major steps in the experiment involved: reviewing and signing an informed consent form, completing a demographic/background questionnaire concerning their experience with land navigation, eye-tracking system

calibration, a familiarization brief covering the simulation system, a practice navigation route, a navigation brief on the trial route, an opportunity for map study followed by the navigation trial. During map study, participants were asked to complete a questionnaire concerning anticipated difficulty of the navigation task. After completion of the navigation trial exercise, participants were asked to provide debrief commentary and finally complete a questionnaire regarding the actual difficulty of the navigation task. A detailed description of these steps follows.

After reading and signing the informed consent form, participants completed a background questionnaire, also listed in Appendix B. The background questionnaire included basic demographic information as well as specific information concerning training and experience related to terrain association – either in aviation or ground-based context. Background demographic and experience information was collected to augment assessment of individual's level of expertise. Following the overview brief and informed consent form, participants were introduced to the simulation equipment and task.

Eye tracking system calibration involved following the FaceLabs standard protocol. Prior to calibration, ambient lights were turned off to allow for consistent lighting. The cameras' position and orientation were adjusted to accommodate the user and provide optimal eye tracking performance. The camera height and distance from the screen were adjusted to optimize the size of the image of the users tracked facial features. If necessary the angle of the cameras was also adjusted to maintain key features within the field of view. If any adjustments were made to the camera's angle relative to the horizon or their position and orientation relative to each other, the new configuration was calibrated in accordance with vendor recommendations. If there was no change to these key configuration parameters, the minimum procedure of re-verifying the calibration was completed. After the camera configuration was verified, the system was calibrated for each user. Calibrating for an individual user followed standard vendor protocol. The calibration protocol included the software system collecting reference picture of the participant and the operator selecting and adjusting reference marks and verifying the tracking accuracy. The participant's reference head model was modified and tracking parameters adjusted until a minimum tracking accuracy of 2.0 degrees was achieved.

Following system calibration, the participants were briefed on the operations of the joystick and on the map and instrument panel display characteristics. A scripted version used in the study is provided in Appendix B. For the practice/equipment familiarization route, waypoint markers (red spheres) were rendered 150' above the navigation waypoints. Participants were briefed that waypoint markers would not be depicted during the navigation exercise portion of the study. The practice route is shown in Figure 12. The practice route is in the same general vicinity, roughly 8 NM to the northwest of the trial route. The map contour interval is the same and the terrain is very similar to the trial navigation route. Before starting the navigation practice route, participants were asked if they had any questions. Participants were then provided the opportunity to fly the navigation route to gain familiarity with the joystick controls, map and instrument panel displays.

Following the calibration phase and equipment familiarization navigation route exercise, participants were briefed on the trial navigation route. The equipment familiarization brief involved a separate PC workstation utilizing FalconView flight planning software system widely employed by diverse communities within DoD. The complete brief is included in Appendix B. The route is depicted in Figure 12. During the map study phase, participants were asked to complete a questionnaire that rated the anticipated difficulty of navigating each leg using terrain association. Participants were provided unlimited time for map study.

After completing the map study phase, participants were directed back to the navigation task simulation. Evaluators then re-verified screen calibration by having users momentarily fixate on the four corners and center of each display area. If calibration appeared to be off significantly, evaluators directed participants through the screen calibration procedure until acceptable criteria were achieved. Once the calibration was verified, participants were reminded that the map display would perform exactly as it did during the previous equipment familiarization exercise. They were also reminded that there would be no waypoint markers visible in the out-the-window view. Evaluators also informed participants to provide any verbal commentary on their navigation performance but that maintaining orientation was the priority over providing verbal commentary. Participants were briefed that occasionally during the exercise evaluators would ask them

to describe their current position. Participants were informed that if their answer didn't represent an accurate solution and they were proceeding too far off course the evaluators would provide directions back to the nearest appropriate waypoint on the navigation route. The decision concerning when to query participants and provide direction back to the route was subjective. The study procedure involved collecting scan data during various phases of navigating. One aspect of particular interest is how navigators recognize when they are off course (that is, when their perceived position no longer matches their actual position) and how they react after they recognize they have made a navigation error. Evaluators decided when to query and provide direction to allow participants maximum opportunity to recognize errors and attempt recovery. The limit on how far evaluators could allow participants to fly was the range of the map display. Since the map did not translate based on aircraft location, evaluators were limited in the distance participants could travel from the pair-wise mean of the waypoints.

F. PARTICIPANTS

Participants were recruited from the Naval Postgraduate School faculty, staff and student body. Recruiting notices were posted on the school's student check in page and bulk email requesting participants were sent to all students. The study was approved by the Naval Postgraduate School Institutional Review Board. Participants were recruited from e-mail advertisement through NPS e-mail account holders. All the participants were given written informed consent to participate, with the right to withdraw at any time.

There were 19 male military personnel, 29 to 40 years of age, and only 15 of them completed the study due to eye-tracking device calibration issues. Only twelve participants flew up to waypoint 5 without experimenter intervention and the following study only analyzed the experimental data between waypoint 2 and waypoint 5 from those twelve participants. This procedure is to remove confounding factors that can be generated from different route difficulties. All of them took overland navigation classes before participation to suffice the minimum skill requirements for the study. Three participants were helicopter flight instructors and two participants had other navigation-related instructing experiences. Total-flight-hours varied from 0 to 3,100 hrs (avg = 1,488 hrs, std = 1,104 hrs) and overland-flight-hours varied from 0 to 2,500 hrs (avg = 612 hrs,

std = 853 hrs). Eight participants are from USN, three from USMC. The remainder of the participants were from the United States Air Force, United States Army and Brazilian Navy. No special neurological, visual acuity, or spatial ability tests were performed.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ANALYSIS AND RESULTS

A. OVERVIEW

The hypothetical model derived in Chapter II.E.1.d drove the following quantitative expected differences in experts and novices: Experts would change views between the map and out-the-window view more frequently. For each scan of the map or out-the-window scene, experts would fixate on fewer features spending less time dwelling on each of these. The model also predicted that expert's dwell times would be more consistent, while novice's dwell times would vary more. Finally, the model predicted that experts would spend proportionally more time scanning out the window compared to time spent viewing the map. Hypothesis testing for these measures is discussed below in 'quantitative analysis.'

Although there are currently no measurable statistics associated with higher level scan characteristics, the model predicts additional traits of expert versus novice scan habits. Overall, the model predicts that experts would have more organized scan patterns. Experts would tend to manage their out-the-window scans to ensure overlap with previous out-the-window scans. Experts would tend to scan the map ahead of the aircraft a consistent distance. The model predicts considerable overlap and consistent time between scanning correlating features in the out-the-window and map views. These are explored using a visualization tool and discussed in more detail in Chapter IV.B qualitative analysis.

B. QUANTITATIVE ANALYSIS

Based on previous research (Sullivan,1998, McLean, 1999, Wright, 2000, Lennerton, 2001, Kulakowski, 2003, & Hahn, 2004), initial regression analysis was conducted using standardized measures of instructing experience and total flight hours as predictor variables. A summary of this analysis is included in Table 5. Significance level for all tests was .10.

To examine any possible interaction effect, an interaction variable was included in follow on analysis. The interaction term, also based on experience from previous studies

(Beilstein, 2003; Hahn, 2005; Kulakowski, 2004; Lennerton, 2004; Sullivan, 1998; Wright, 2000), was derived by assigning a value of one for individuals with no instructing experience, two for individuals with ground-based terrain navigation instructing experience and three for individuals with aviation terrain navigation training experience. Similarly, because flight hours were skewed, a value of one was assigned for individuals with no flight time, two for individuals with between zero and 1500 hours, three for between 1500 and 2350 hours and four for over 2350 hours. The interaction variable was the product of these two values. A summary of the second regression analysis is included in Table 7.

1. Global Summary of Results

There were significant results for seven of the sixteen dependent measures when the model included just two predictors. When the interaction term was included in the model, significant results were found for an additional five dependent measures. Combined, there were significant results for twelve of sixteen measures.

Dependent Measure	Expected Result	Significant Results, Source	Comments
Overt Performance Measures			
Flight path RMS error	RMS is a poor indicator of expertise.	No	Lack of significance tends to support predicted model.
Actual versus Ideal Flight Time	Experts will come closer to arriving on time.	Primary	Supports predicted model
Basic Dwell Characteristics			
Dwell duration – median	Experts' median dwell will be less than novices.	Primary	Supports predicted model
OTW dwell – median	Experts' median OTW dwell will be less than novices.	Primary	Supports predicted model
MAP dwell – median	Experts' median map dwell will be less than novices.	N/A	Data Not Normally Distributed
Dwell duration – mean	Experts' mean dwell will be less than novices.	Secondary	Supports predicted model
OTW dwell – mean	Experts' mean OTW dwell will be less than novices.	N/A	Residual Not Normally Distributed
MAP dwell – mean	Experts' mean map dwell will be less	Secondary	Supports

	than novices.		predicted model
Dwell duration – STD	STD of expert dwell duration will be lower than novices.	Secondary	Supports predicted model
OTW dwell – STD	STD of expert OTW dwell duration will be lower than novices.	Secondary	Supports predicted model
MAP dwell – STD	STD of expert map dwell duration will be lower than novices.	Secondary	Supports predicted model
Higher-level Scan Characteristics			
Percentage of flight time in Dwell	Experts will have more saccade time, thus shorter overall dwell time.	Primary	Supports predicted model
Fixations per OTW view	In each OTW view, experts will fixate on fewer points	Primary	Supports predicted model
Fixations per MAP view	In each map view, experts will fixate on fewer points	N/A	Data Not Normally Distributed
View Changes/Flight Time	Experts will change views more frequently than novices	Primary	Supports predicted model
OTW Dwell / Total Dwell	Experts will spend more of their dwell on OTW than map	Primary	Contradicts predicted model

Table 7. Summary of results

Two variables failed normal distribution assumption checks: the median map dwell time and the number of fixations in each map scan. For the dependent measure of mean out-the-window dwell time, results from the first analysis were not significant and the residuals were not normally distributed in the second analysis. Thus, no conclusions could be drawn regarding mean out-the-window dwell time. For the dependent measure of root mean square (RMS) flight path error, results were not significant in either analysis. Significance for RMS flight path error would have tended to contradict the predicted model. For the twelve dependent measures with significant results, the analysis supported the hypothetical model for eleven dependent measures. There is strong supporting evidence that the variable which contradicted predictions was affected by a floor effect. These results are discussed in detail below. Based on these results, expertise can be predicted with a high degree of accuracy using eye scan information alone.

Variable (z-score units)	Intercept	Instructor Experience	Flight Time	Interaction Term	F	Adjusted R^2
Dwell duration - mean	6.94E-10(0.194)	2.314(0.684)***	0.919(0.469)*	-2.775(0.844)**	5.466**	0.549
MAP dwell -mean	6.90E-10(0.209)	2.312(0.737)**	0.853(0.505)	-2.426(0.909)**	4.341**	0.477
Dwell duration STD	8.94E-10(0.155)	2.799(0.547)****	1.265(0.375)***	-3.242(0.675)***	10.068***	0.712
OTW dwell duration STD	6.55E-10(0.246)	1.986(0.868)*	1.039(0.595)	-2.619(1.071)**	2.387	0.274
Map dwell duration STD	6.79E-10(0.185)	2.330(0.654)***	0.757(0.448)	-2.382(0.806)**	6.249**	0.589

* p<.10, ** p < .05, *** p < .01, ****p < 001

Table 8. Secondary Analysis: Includes Predictor Variables of Instructor Experience, Flight Time and an Interaction Term

2. Raw Performance Variables.

A fundamental assumption behind this research is that raw indicators of performance, including distance from a straight-line path between waypoints, can be an inaccurate indicator of expertise. Regression analysis tended to be consistent with our assumption, as neither instructing experience, flight hours, nor the interaction were significant predictors of RMS distance between the actual and ideal flight path. A strong relationship between predictor variables and raw performance measures such as RMS error would have tended to suggest that performance measures alone could be sufficient for assessing expertise. Although this finding does not directly support the hypotheses, correlation between expertise and RMS error would have tended to contradict our hypothesis.

Dependent Measure	mean	std
RMS error	8.72 ft	5.71 ft
FLT error	4.17 sec	5.12 sec
Median dwell duration	231 msec	48.2 msec
Median OTW dwell duration	231 msec	41.4 msec
Median MAP dwell duration	271 msec	132 msec
Mean dwell duration	371 msec	82.1 msec
Mean OTW dwell duration	337 msec	80.6 msec
Mean MAP dwell duration	454 msec	184 msec
STD dwell duration	411 msec	113 msec
STD OTW dwell duration	359 msec	123 msec
STD MAP dwell duration	481 msec	201 msec

Scan time in flight time	87.4 %	5.10 %
Num. of fixation points	262	51.0
Num. of OTW-MAP view changes	123	61.2
Num. of fixations per OTW view	3.33	1.81
Num. of fixations per MAP view	1.76	0.806
OTW scanning time	57.7 %	10.4 %

Table 9. Descriptive statistics for dependent measures

As expected, the difference between actual and ideal flight time was predicted by expertise. Using the initial analysis method, flight hours were a significant predictor (adjusted R squared = 0.373, p = 0.053) of absolute value of actual versus ideal flight time (un-standardized bft = 0.003, SEft = 0.001.) Since experts may intentionally deviate from course, RMS error was not predicted to be strongly correlated with expertise. However, based on the assertion that expert pilots will be able to compensate for any difference in time these excursions will cause, the hypothetical model predicted that experts would arrive at checkpoints nearer to the ideal time. In combination, these measures suggest that while expert navigators may deviate from the planned course, they tend to meet timing constraints.

3. Basic Dwell Characteristics

The hypothetical model predicts that for map, out-the-window and total time, experts will have lower mean and median dwell times with a lower standard deviation. Of these nine parameters (mean, median, and standard deviation of dwell time in out-the-window, map and total views), there were significant results and the analysis supported our hypothetical model for seven variables. Median dwell data for map scans was not normally distributed, so regression analysis of median dwell time could not be performed. For mean of out-the-window dwells, no significant results were found in the initial analysis. In the follow on regression analysis that included the interaction variable the residual was not normally distributed. Since distribution of the residual violates regression analysis assumptions, no conclusions could be drawn regarding the mean of out-the-window dwell time.

For the median of overall dwell time, results from the initial analysis were significant and supported the hypothetical model. Flight time was a significant predictor

(adjusted R squared = .311, p = 0.027) of the median of overall dwell duration (un-standardized bft = -0.029, SEft = 0.011.) Thus, for every 100 flight hours, on average the model predicts that median of overall dwell will decrease 2.9 msec. Similarly, the first analysis returned significant results for the median of OTW dwell measurements. Flight time was again a significant predictor (adjusted R squared = 0.379, p = 0.019) of median OTW dwell (un-standardized bft = -0.026, Seft = 0.009.) For every additional 100 flight hours, on average individual's median OTW dwell time decreased 2.6 msec. Median map dwell times were not normally distributed, so no conclusions could be drawn concerning this measure.

The mean of both overall and map dwell duration matched the hypothetical model. For OTW dwell, results were not significant for the first analysis. In the second analysis, the residuals were not normally distributed. Thus no conclusions could be draw regarding mean of OTW dwell. The interaction term of expertise rating was a significant predictor (adjusted R squared = 0.549, p =0.011) of the mean of overall dwell duration. Likewise the interaction term was a significant predictor (adjusted R squared = 0.477, p = 0.028) of the mean map dwell duration.

The prediction of the hypothetical model that expert's dwell times would vary less than novice's was also supported. For overall, OTW and map standard deviation of dwell times, the first analysis returned no significant results. The secondary analysis returned significant results for each of these dependent measures. The interaction term was a significant predictor of the standard deviation of overall dwell time (adjusted R squared = 0.712, p = 0.001.) The interaction term was also a significant predictor of both the standard deviation of OTW dwells (adjusted R squared = 0.274, p = 0.040) and the standard deviation of map dwells (adjusted R squared = 0.589, p = 0.018.)

4. Scan Management Characteristics

The analysis also supported predictions for some, but not all of the parameters associated with more deliberate scan control. The percentage of time spent in dwell, the number of fixations per out-the-window view, and the view changes per unit time matched predictions. Data associated with the number of map dwells per view was not normally distributed, so no conclusions could be drawn for this dependent measure.

Significant results that contradicted hypothetical model expectations were found for the percentage of dwell time spent out the window versus on the map. The contradiction between expected and measured results for percentage of dwell time spent in the out-the-window view could be explained by a possible floor effect. Additional support for the floor effect is included in Chapter IV.A.

Based on the assertion that expert navigators will change views more frequently; the model predicted that experts would spend a smaller proportion of their flight time in dwell with more time spent in saccade and transition. This assertion was supported by the initial analysis. Flight time was a significant predictor (adjusted R squared = 0.295, p = 0.031) of total time spent in dwell (un-standardized bft = -0.003, SEft = 0.001.) Analysis of total time spent in dwell supports the model's prediction that expert are better able to manage frequent changes from map to out-the-window view and require less fixation time to gather information.

The hypothetical model predicted that experts will dwell on fewer features during each out-the-window view. Based on the primary analysis, flight time was a significant predictor (adjusted R squared = 0.264, p = 0.055) of the number of fixations during in each out-the-window view (un-standardized b_{ft} = -0.001, SE_{ft} < .001.) Thus as flight hours increase, the number of features that individuals dwell on in the out-the-window view decreases. As the model predicted, experts changed view per unit time more frequently than novices. The total number of view changes over the route varied significantly (adjusted R squared = 0.198, p = 0.058) based on flight hours (un-standardized b_{ft} < 0.001, SE_{ft} < 0.001).

The only result that was contrary to predictions was the percentage of scan time spent scanning the out-the-window view. The model predicted that experts would spend proportionally more time scanning out the window. Analysis indicated statistically significant results (adjusted R squared = 0.216, p = 0.052) in direct contradiction to this assumption (un-standardized bft = -0.006, SEft = 0.003). That is, experts on average spent proportionally less time looking out the window. For every 100 flight hours, on average individuals spent 0.6 percent less of total scan time in out-the-window dwell. One possible explanation is that the task resulted in a floor effect. It is feasible that,

particularly for experts, the task was not overly taxing. This is likely considering that navigation is normally one of several if not many other tasks being conducted concurrently. In absence of additional tasks commensurate with real-world conditions, experts may have spent much more time reviewing the map than they normally would. Given that navigation was expressly briefed as the point of the study and several expert pilots completed the route with relative ease, this seems highly probable. This possibility is supported by visual analysis of scan patterns and is discussed in more detail in Chapter IV.A.

While there is no clear indication of why statistically significant results were found for the number of map fixations, a scatter plot Figure 18 demonstrates that a single outlier may have skewed the data.

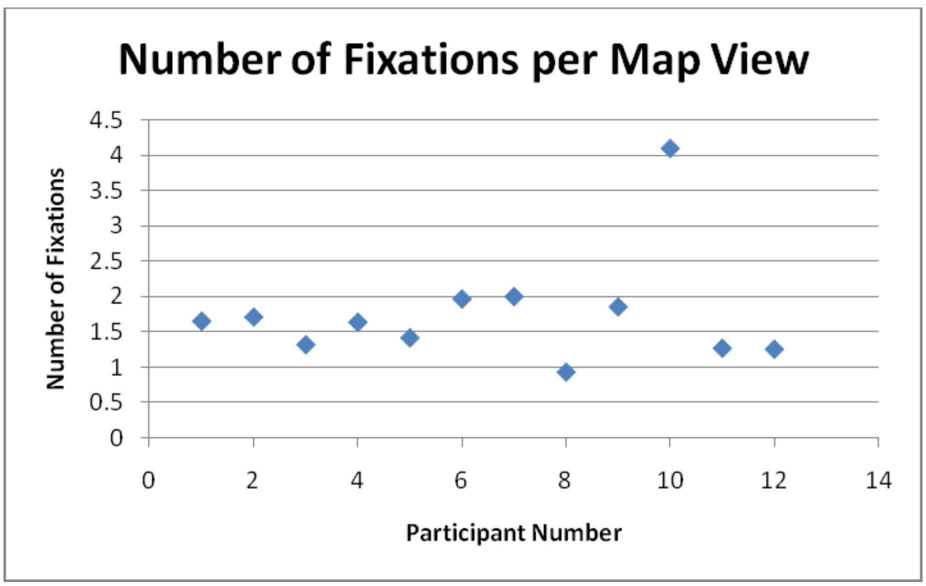

Figure 18. Number of Fixations Per Map View

Participant 11's fixation points per map view were several standard deviations away from the mean. The mean value was 1.7 fixations per dwell with at standard deviation of .8. Participant 11's average number of fixations/map view was nearly 3 standard deviations from the mean.

Variable (z-score units)	Intercept	Instructor Experience		Flight Time		F		Adjusted R^2
Flight Path RMS Error	2.76E-17(0.304)	-0.279(0.320)		-0.092(0.320)		0.460		-0.109
Flight Time Error	8.33E-11(0.229)	-0.516(0.241)	*	0.536(0.241)	*	4.272	**	0.373
Percent Dwell Time	-8.33E-11(0.242)	0.147(0.255)		-0.651(0.255)	**	3.296	*	0.295
F/D*100	8.33E-11(0.256)	0.088(0.269)		-0.603(0.269)	*	2.514		0.216
MAP scan duration (sec)	-8.33E-11(0.252)	-0.183(0.265)		0.608(0.265)	**	2.717		0.238
H/D*100	-8.33E-11(0.256)	-0.088(0.269)		0.603(0.269)	*	2.514		0.216
Dwell duration - median (msec)	-8.33E-11(0.240)	0.127(0.252)		-0.664(0.252)	**	3.484	*	0.311
OTW dwell - median (msec)	-8.33E-11(0.227)	-0.110(0.239)		-0.680(0.239)	**	4.356	**	0.379
View Changes (numbers/sec)	-8.33E-11(0.259)	-0.032(0.272)		0.589(0.272)	*	2.356		0.198
Num. of fixations per OTW view	8.33E-11(0.248)	0.338(0.261)		-0.575(0.261)	*	2.977		0.264
* p<.10, ** p < .05, *** p < .01, ****p < 001								

Table 10. Relationship of Instructor Experience and Flight Hours on Scan Characteristics

Spearman correlation, shown in Table 11, provided additional support for the hypothesis. Applying Spearman correlation, we found that readily observable performance metrics of RMS Error and Flight Time Error are positively correlated with each other ($\rho = .664$, $p = .022$). Also as predicted, many of eye-tracking metrics are positively correlated with each other. Of note, neither of the flight performance metrics, RMS Error and Flight Time Error, were correlated with any of the eye-tracking metrics. The lack of correlation between overt measures (RMS and Flight Time Error) and eye-tracking measure suggest that raw performance measures are independent of measures of underlying cognitive processes. Thus, Spearman correlation supports our hypothesis that exclusively relying on flight performance may not provide a reliable indicator of proficiency.

	RMS error	FLT error	Median dwell	Median OTW dwell	Median MAP dwell	STD dwell	STD OTW dwell	STD MAP dwell	Percentage scan in flight time	Num. of fixation points	OTW-MAP view changes	Num. of fixations per OTW	Num. of fixations per MAP
RMS error	-												
FLT error	.66**	-											
Median dwell	.20	.08	-										
Median OTW dwell	.30	-.04	.89****	-									
Median MAP dwell	-.01	.08	.88****	.59**	-								
STD dwell	.13	.05	.80***	.57*	.85*	-							
STD OTW dwell	-.04	-.17	.61**	.56*	.51*	.75*	-						
STD MAP dwell	.20	-.06	.54*	.39	.66*	.71**	.21	-					
Percentage scan in flight time	.01	-.17	.91****	.77*	.91****	.86****	.65**	.71**	-				
Num. of fixation points	-.24	.06	-.93****	-.86****	-.79*	-.86****	-.61**	.69**	-.91****	-			
OTW-MAP view changes	.19	.37	-.64**	-.57*	-.66*	-.55**	-.43	-.45	-.76***	.65**	-		
Num. of fixations per OTW	-.20	-.31	.39	.33	.52*	.34	.25	.45	.58*	-.34	-.85****	-	
Num. of fixations per MAP	-.08	-.09	.07	-.03	.13	-.04	.06	-.18	.10	.03	-.59**	.59**	-
OTW scanning percentage	-.01	-.15	.61**	.72*	.51*	.41	.52*	.34	.62**	-.51*	-.45	.55*	-.11

* p<.10, ** p < .05, *** p < .01, **** p < .001

Table 11. Spearman Correlation Data

C. QUALITATIVE ANALYSIS

In addition to the quantitative analysis, several versions of a visualization tool were developed and explored to help understand the statistical analysis or provide additional insight into novice and expert scan patterns. These visualization tools have provided valuable insight and inspired additional metrics for future evaluation. While these metrics are not yet conclusive predictors of expertise, they represent a promising area of future study. Visualization of scan patterns also remains a promising are for future discovery.

1. Sample Qualitative Analysis and Exploration of Additional Predictor Variables

The analysis tool described in Chapter II.D.2 provided insight into how expert scan patterns are organized. This lead to preliminary development and evaluation of variables to more accurately capture some the higher-level attentional controls exhibited by experts. Table 2 depicts expected results for expert performance. Based on the original CTA and SME insights from previous studies (Beilstein, 2003; Kulakowski, 2004; Lennerton, 2004) experts would be expected to consistently scan the map well ahead of the aircraft's position. The model also predicts a high degree of overlap among features scanned in the terrain and subsequently scanned on the map within a given time frame. This pattern was demonstrated frequently by expert performers. Visual analysis indicated that novices demonstrated this pattern less frequently.

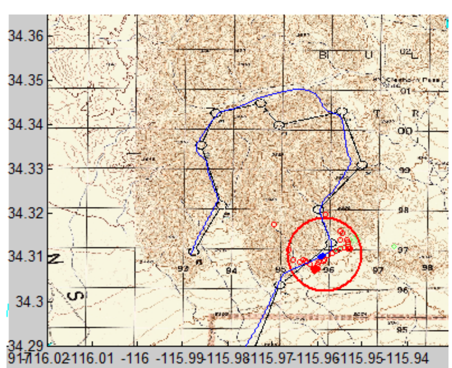

Figure 19. Participant map scan behind, as well as ahead of current position. Blue—aircraft track, red—map scan points

Visual analysis provided insight into performance and also revealed patterns that were not expected. At least one expert followed a very consistent pattern that involved spending substantial time scanning the map behind the aircraft's current position. This can be seen clearly in Figure 19. The red circles immediately behind the aircraft position represent recent map scans. This pattern is repeated consistently from waypoint one through six. Based on visualization of scan and debrief comments, this appears to be a deliberate and helpful strategy as noted from a description of navigation past waypoint six. Almost immediately after making the left turn at waypoint six, the participant perceived that he had made the turn early and was south of course. He was, in fact, only slightly north of course. The participant's small deviation from the expected position was sufficient to cause confusion between his current view and similar map depiction of terrain immediately south of his position. He immediately turned right nearly 90 degrees to correct. Soon after making the turn, he entered a valley perpendicular to the flight path and wider than expected. He turned left to follow the valley with the intent of disambiguating between the two east-west oriented valleys. Shortly after making this turn he identified a unique terrain formation in the floor of the valley. The unique terrain formation was only present in the northernmost valley. His next map scan he located the formation on map. Correlating this key feature allowed the participant to recover from his perceived error and continue successful navigation. With only raw performance data such as RMS error or timing data, an observer would have no way of knowing if the course deviations were intentional. Without scan data, there would be no way at any point in that evolution to know if recovery was likely and if effective navigation and learning was taking place

Even after the subject had re-oriented himself and was successfully navigating along course, his tendency was to focus back to the area of uncertainty. In at least one case, an expert appeared to be drawn to dedicate any additional mental resources to map study in an attempt to solve the puzzle of how the mismatch might have occurred. This observation tends to support the earlier assertion that the division of scan between the out-the-window and map view was impacted by a floor effect. It seems highly likely that if the navigation or other combined tasks were beyond moderately challenging, individuals would have dedicated less time to studying the map. While this analysis is

highly anecdotal and difficult to substantiate, post-event debrief comments strongly support this premise. Additional studies to evaluate are highly encouraging.

The deliberate strategy of checking the map immediately behind the presumed aircraft position may have played a role in the participant's ability to recover. From experts' debrief comments; it appears that they tend to maintain multiple possible navigation solutions simultaneously. They continually challenge these solutions against the widest set of evidence they can gather—to include verifying the terrain they recently covered does is in fact correspond to terrain represented on the map. Checking recently covered terrain against the map depiction could provide an opportunity for hypothesis confirmation and alternative hypothesis generation. In the example above, the participants debrief comments indicated that when faced with unexpected terrain features in view, their alternative possible location was based on map scan in the vicinity of recently-covered terrain. Although the participant was mistaken in the assumption that he was off course, he could not have mismatched terrain in view with a plausible alternative location if he had been not been comparing terrain covered against multiple possible locations on the map. Without robust hypothesis generation and testing, including terrain recently covered, it would not have been possible for the participant to identify the point at which he thought he had made the turn early. Of the top three expert navigators in this study, two made minor errors and corrected themselves, each described similar structures for recovery in debrief comments.

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

A. FINDINGS AND IMPLICATIONS

Based on the analysis provided above, this work demonstrates that neurophysiological markers can be used to indicate aspects of a trainee's cognitive processes that are useful for cuing an instructional system. For the complex cognitive task of helicopter overland navigation via terrain association, the fundamental characteristics of fixation and dwell time can reliably distinguish between experts and novices. In addition, visualization of scan pattern is useful for informing instructors of trainee's strategy and reveals unexpected strategies of experts. The implications of these findings in the context of earlier studies follow.

Neuromarkers to improve cognitive task analysis. Cognitive task analysis, particularly for complex cognitive tasks is labor intensive and produce widely variable outcomes (Ericsson, 2006) (Crandall, Klein, & Hoffman, 2006). This is due in part to the fact that experts may have a difficult time expressing how they achieve outstanding results (Barnard, 2000). Because they are relying on a high level of automaticity and can process at a nearly subconscious level, describing their own strategy can be extremely difficult (J. K. Phillips, Shafer, Ross, Cox, & Shadrick, 2006). As seen in the example earlier, neuromarkers can provide information on underlying strategies experts employ. In the case of eye tracking for navigation tasks, data can be collected in real time without interference between the user and the environment. The only impact is the time required to calibrate the equipment. Passive collection of information on internal processes associated with skill acquisition provides information that is not available to human observers. This information can reveal strategies that the operator may not be aware of or be able to describe. This has two important implications. First, tracking cognitive processes associated with development of expertise could be used to assess trainees and guide instruction. Second, tracking could be used to elucidate strategies that could then be incorporated into existing training regimes that may not involve neuromarker tracking.

Neuromarkers to improve the design and evaluation of training. Based on the work related to training design tailored to level of expertise (Anderson, 1981; Bilalic,

McLeod, & Gobet, 2008; Charness & Tuffiash, 2008; Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 1999; Liu, Yuan, Fan, Liu, & Kang, 2009; Sweller, 1988; J. J. G. van Merriënboer & Sweller, 2005) described in Chapter II.0, training can be improved dramatically when it properly assesses and accounts for trainee's level of expertise and cognitive load. Adaptive training has tremendous potential; however the current set of cues on which adaptation should be based is extremely sparse (Coyne & Baldwin, 2003). This work has established that eye scan can provide a valuable cue for guiding the instructional process. Chapter II.C outlined the downside of relying on performance comparisons for training effectiveness evaluation. Not only does performance vary as expertise is developed, deviation from performance metrics normally provides little diagnostic value. Tracking expertise via underlying mechanisms such as scan characteristics and patterns may provide a more accurate representation of actual skill and may also provide useful diagnostic information, such as why performance metrics are not being met.

Neuromarkers to improve models of expertise. Currently models of expertise are built and connections made to underlying information processing models primarily via human measurable events including tests, interview and observation (J. K. Phillips, Shafer, Ross, Cox, & Shadrick, 2006) (Ross, Phillips, Klein, & Cohn, 2005). This work has demonstrated that cues that are not human observable can be used to validate or extend our understanding of expertise and our models of information processing.

Neuromarker selection process. The current process described in this work for identifying neuromarkers for complex cognitive tasks relied heavily on a series of interconnected research projects. Each of these relied heavily on subject matter experts to guide the research project. Although the selection of neuromarkers is supported by referencing the cognitive task analysis for this study, further work to codify the process for identifying salient neuromarkers is warranted. For example, the CTA listed in Appendix A and the description of navigation performance provided in Chapter II.E.1.a, highlight the importance of confidence. Previous work (Schmorrow, Estabrooke, Grootjen, Campbell et al., 2009) has demonstrated that EEG signals may be used to reliably detect when participants are guessing or certain during static recognition tasks. Although EEG measures are a more direct measure of cognitive activity than eye

tracking, collecting EEG information is much more intrusive. It is also more difficult to meaningfully interpret EEG signals on dynamic, complex tasks. Establishing methods for classifying the ease-of-use, detect-ability and diagnostic value of various neurophysiologic signals and their relationship to cognition would improve the process of connecting neurophysiologic and instructional systems.

Description of cognitive factors related to skill acquisition and means to reliably identify corresponding markers. By definition, complex cognitive tasks have multiple inter-related components that can be addressed using a wide variety of strategies. Since operators must attend to multiple environmental cues and can apply a wide range of strategies, it can be very difficult to break down a task into factors that relate to detectable differences in underlying cognition across stages of expertise. The approach followed in this research process was successful in translating task elements from a CTA into a high level description of strategies and novice/expert differences that could be further translated to measurable differences detected through eye-tracking.

B. LIMITATIONS OF THE CURRENT APPROACH

As promising as the results appear, some caution is warranted. One area of concern is the process for selecting tasks and appropriate neurophysiological markers. The approach followed in this work was influenced by numerous intermediate studies, each of which incorporated extremely close coupling of multi-domain technical experts alongside subject matter experts. Although selection of eye tracking is supported by the CTA, the degree of influence of previous studies is unknown. Generalizing the process for selecting neuormarkers across other problem domains remains open for further investigation. In addition, numerous decision related to experimental design should be addressed in follow on research. While it was appropriate in this early investigation to limit task complexity, each compromise to support data collection and analysis increases the level of abstraction from the real world, thus making conclusions about transfer to the real world more difficult. Recommendations for further research are covered in more detail in the following section; however, the following topics are briefly presented for this task and for generalizing this process.

1. Refining the Current Study

The following paragraphs outline methods to refine and extend the current study:

Allow participants to plan their own route. Navigation performance is normally a function of both planning and execution. In the study presented here, individuals did not have the opportunity to plan their own routes. Presumably, experts could plan routes more efficiently to take advantage of key features. Improved planning would result in a simplified task during in flight navigation and may have lead to important differences in individual performance. Unfortunately, without standardized routes it would have been extremely difficult to make comparisons across subjects.

Minimize time and improve accuracy of calibration. Eye tracking procedures were labor intensive and did not always produce accurate results. For two of the nineteen recruited participants, eye tracking equipment could not be calibrated with sufficient accuracy. Options for experimental design are limited based on the time participants can contribute. Eye calibration procedures took up to 15 minutes of the 75-90 minute session. The time to calibrate eye tracking measures for each user would have been much substantially longer if the full field of view and normal cockpit interface was provided.

Increase field of view commensurate with the aircraft. In the context of this exploratory research, we were able to meet the research objectives using a simulation that provided a limited field of view for the out-the-window scene compared to the field of view available in the real environment. The statistical and qualitative analysis supported the hypothesis that experts would select more appropriate features in the out-the-window view faster and would spend less dwell time on these features. With increased field of view we would expect to see a more pronounced effect. Giving a novice more information by providing a wider field of view may further complicate the process of selecting cues and increase the time required to select features of interest.

Provide more authentic map controls. Providing paper maps exactly as they are used in the cockpit would have prevented meaningful data collection on map gaze. Providing a means for participants to rotate the map manually would have been more realistic; however with current technology this would have required additional training time, lengthened the period of study and created additional mental workload that could

have varied by individual. Rotating the map automatically is not an ideal solution. Although it is doctrine to orient the map in the direction of travel, this is normally a manual process.

Provide more realistic tasking. As noted in Wright (2002), navigation is always a means to an end. It is never the exclusive mission. The close air support mission outlined in (King & Lakey., 2006) would provide a more realistic overall context. As discussed in the analysis section, participants had more free time to dedicate exclusively to one component task in the simulation than the ever would in any aircraft setting.

Evaluate tasks in a more realistic cockpit environment. As noted by (Lennerton, 2004), transfer issues and real world performance issues are always influenced by the physical constraints and denial of information associated with the operational environment. In all likelihood studies in more realistic environments would lead to more pronounced differences between novices and experts. Individual differences across multiple training platforms could provide novel insight on the interaction of training media, simulation fidelity, expertise and training tasks.

2. Generalizing This Approach

While this work provides promising leads in one specific domain, the ultimate effort is to extend this work across multiple domains exploiting multiple neuromarkers. Therefore the following topics would need to be addressed:

Neuromarker selection process. The current process described in this work for identifying neuromarkers for complex cognitive tasks relied heavily on a series of interconnected research projects. Each of these relied heavily on subject matter experts to guide the research project. Although the selection of neuromarkers is supported by referencing the cognitive task analysis for this study, further work to codify the process for identifying salient neuromarkers is warranted. While the CTA provides a basis for validating candidate neuromarkers, it seems likely that the process should start from the view of experienced practitioners and instructors.

This is particularly appealing in the context of combining markers. On more complex tasks, it may be important to know what component of the task the trainee is

attending to and how certain or comfortable they are with their solution. In this case an eye scan system could cue an EEG system. In combination, these systems could indicate the degree of certainty trainees had when a correlating feature is scanned on both the map and out-the-window view. While this information may be able to be verified in a comprehensive CTA, starting from the perspective of the trainer observing and making inferences about trainee's cognition seems likely to provide a broader and more reliable set of candidate markers.

Thorough descriptions of cognitive factors associated with acquisition of skill and means to reliably identify corresponding markers. By definition, complex cognitive tasks have multiple inter-related components that can be addressed using a wide variety of strategies. Involved complex cognitive tasks will have multiple aspects of 'state' that will take precedent during different phases of execution. In this study a single complex cognitive task was identified, isolated in a tailored simulation and task and studied. For this task it was relatively straightforward to correlate attention management as a key indicator of cognitive ability and thus relative expertise. Likewise identifying scan pattern as an indicator of attention management is straightforward. It was supported by SME opinion, previous studies (Beilstein, 2003; Lennerton, 2004; Sullivan, 1998; Wright, 2000) and current eye scan research (Bellenkes, Wickens, & Kramer, 1997); (Liu, Yuan, Fan, Liu, & Kang, 2009); (Marshall, 2007).

To extend this work across other domains and consider other markers that could indicate useful insight on trainee's cognitive processes, investigators would need to standardize the methods for indentifying features of interest. While the CTA may be useful for validating the potential salience of signals related to internal processes, identifying cognitive factors associated with learning should be viewed from an instructional perspective.

Exploring cases where indicators of cognition contradict expected results. In this example study, predictions for only one of fourteen dependent measures did not match expectations. As discussed in Chapter IV.B.3, experts spent comparatively less time looking out the window than novices. The model had predicted the opposite: experts were expected to spend more time using the out-the-window view. The

difference between the model prediction and observed results is likely due to an artifact associated with the experimental design. However, the difference between model prediction and observed results raises the issue of providing validation of predicted results. There are two sides to the question of validating results: What should happen when markers contradict expected results? Are there methods to verify that the markers and metrics selected do in fact provide the expected insight into the targeted aspects of operator cognition? In this study example, the path for investigating contradictory results seems plausible. The investigation of plausible explanations started with a comparison of the expected task and the actual task. In this case, the important difference between the actual and expected task was the amount of task loading. Differences in scan distribution could easily be attributed to this. The issue of verification that markers selected a

Cuing instructional interventions. The ultimate aim of this research is to contribute to improvements in the training design and evaluation processes. The point of selecting signals that can provide reliable diagnostic information regarding trainee's cognition is to cue instruction. Raw statistics behavior statistics associated with dwell characteristics can likely indicate when a trainee is lined up to succeed (i.e., their scan characteristics match those of an expert) or when they are at or approaching task overload. Thus, dwell characteristics may be useful for adjusting the overall task difficulty. An adaptive training system could use this information to adjust the level of task difficulty or the level of job aiding provided. This could include automated direct feedback on task performance knowledge of results to build trainee confidence.

Although identifying a trainee's navigation strategy could feasibly be automated at some point in the future; currently, identification of strategy likely requires subject matter expertise and could not be easily automated. Strategy could vary widely based on terrain type, the operator's level of confidence and personal preference. As with expert performance on any complex cognitive task, describing the 'right' approach is highly subjective and will vary widely across top-performing individuals. The process of automating analysis and identification of strategy is further complicated by difficulty associated with automatically assessing contour map representation of features. For example, defining what constitutes a 'salient' feature within a given section of terrain is far from trivial task.

Providing scan visualization as feedback to the student could; however, be valuable to an instructor for their assessment. Scan visualization provides provide an entire level of information not otherwise available. Additionally, providing scan characteristic feedback directly to the student might itself be useful. Providing student's data on their scan efficiency metrics could serve as a sort of second tier of knowledge of results. In the case of navigation, raw knowledge of results would just echo performance metrics such as RMS error and route timing. Providing visual feedback in an after action replay; particularly if compared with one or more expert models could allow the trainee to perform highly effective self-assessment of strategy selection. Currently there is no way for a trainee to replay a navigation event that would allow him to see *why* he may have strayed from course. Since scan visualization is useful for experienced individuals to assess strategy, perhaps providing scan visualization to trainees would allow them to assess their own ability on critical components such as selecting relevant cues.

VI. CONCLUSION AND FUTURE WORK

The design and evaluation of individualized training systems can be improved by identifying neuormarkers that indicate differences in cognition associated with the development of expertise. Subject matter experts can provide insight into neuromarkers that provide insight into when particular neuromarkers could provide useful information. Existing cognitive task analysis product support evaluation of neuromarkers for aiding instruction. When applied to the domain of helicopter overland navigation, eye scan provides useful insight into instruction. Raw performance data can be used to distinguish novices and experts. Visualization of scan data can aid in assessment of strategy and can highlight strategies that are not apparent to experts.

Based on these results and the conclusions draw, the following areas of future research are recommended:

- Generalize the process of identifying complex cognitive tasks and associated neurophysiological markers that reliably indicate cognitive processes associated with acquisition of skill useful for cuing instructional systems.
- Investigate the process for identifying, administering and monitoring the result of instructional interventions applied based on neuromarkers.
- Investigate the impact of improved individualized instruction on training effectiveness.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Anderson, J. R. (1981). *Cognitive skills and their acquisition.* Paper presented at the Carnegie Mellon Symposia on Cognition Series, Hillsdale, N.J.

Barnard, P. J. (2000). Towards a Theory-Based Form of Cognitive Task Analysis. In J. M. Schraagen, S. Chipman & V. Shalin (Eds.), *Cogntive task analysis*. Mahway, N.J: L. Erlbaum Associates.

Beilstein, D. (2003). *Visual Simulation of Night Vision Goggle Imagery in a Chromakeyed, Augmented Virtual Environment.* Naval Postgraduate School.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual Scanning and Pilot Expertise: The Role of Attentional Flexibility and Mental Model Development. *Aviation space and environmental medicine., 68*(7), 569.

Bilalic, M., McLeod, P., & Gobet, F. (2008). Expert and novice problem solving strategies in chess: Sixty years of citing de Groot (1946). *Thinking &; Reasoning, 14*(4), 395-408.

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher, 13*(6), 4-16.

Briggs, L. J., & Ackerman, A. S. (1977). *Instructional design : principles and applications*. Englewood Cliffs, N.J.: Educational Technology Publications.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332.

Charness, N., & Tuffiash, M. (2008). The Role of Expertise Research and Human Factors in Capturing, Explaining, and Producing Superior Performance. *Human factors : the journal of the Human Factors Society., 50*(3), 427.

Chase, W. G. (1982). *Spatial Representations of Taxi Drivers*. Ft. Belvoir: Defense Technical Information Center.

Chase, W. G., & Chi, M. T. H. (1979). *Cognitive skill : implications for spatial skill in large-scale environments*. [Pittsburgh, PA]: University of Pittsburgh, Learning Research and Development Center.

Chase, W. G., & Ericsson, K. A. (1982). *Skill and Working Memory*. Ft. Belvoir: Defense Technical Information Center.

Chase, W. G., Lyon, D. R., & Ericsson, K. A. (1979). *Individual Differences in Memory Span*. Ft. Belvoir: Defense Technical Information Center.

Clark, R. (2003). *Building expertise: cognitive methods for training and performance improvement*. Washington, D.C.: International Society for Performance Improvement.

Clark, R., Nguyen, F., & Sweller, J. (2005). *Efficiency in Learning: Evidence-Based Guidelines to Manage Cognitive Load*: Pfeiffer.

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research, 53*(4), 445-459.

Coyne, J. T., & Baldwin, C. L. (2003). Comparison of the P300 and Other Workload Assessment Techniques in a Simulated Flight Task. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 47*, 601-605.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds : a practitioner's guide to cognitive task analysis*. Cambridge, Mass.: MIT Press.

Di Nocera, F., Camilli, M., & Terenzi, M. (2006). Using the Distribution of Eye Fixations to Assess Pilots' Mental Workload. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 50*, 63-65.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload. *Journal of Cognitive Engineering and Decision Making, 1*, 271-285.

Dick, W., Carey, L., & Carey, J. O. (2007). *The systematic design of instruction*. Boston; MA: Pearson, Allyn and Bacon.

Dreyfus, H. L., & Dreyfus, S. E. (1984). From Socrates to expert systems : The limits of calculative rationality. *Technology in Society, 6*(3), 217-233.

Dreyfus, H. L., Dreyfus, S. E., & Athanasiou, T. (1986). *Mind over machine : the power of human intuition and expertise in the era of the computer*. New York: Free Press.

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers, 34*(4), 455-470.

Ericsson, K. A. (1985). *Skilled Memory*. Ft. Belvoir: Defense Technical Information Center.

Ericsson, K. A. (1988). Skilled Memory and Expertise: Mechanisms of Exceptional Performance. from http://handle.dtic.mil/100.2/ADA193829

Ericsson, K. A. (2006). *The Cambridge handbook of expertise and expert performance*. Cambridge; New York: Cambridge University Press.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review, 100*(3), 363-406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273.

Gagné, R. M. (2005). *Principles of instructional design*. Belmont, CA: Thomson/Wadsworth.

Hahn, M. E. (2005). *Implementation and analysis of the Chromakey Augmented Virtual Environment (ChrAVE) version 3.0 and Virtual Environment Helicopter (VEHELO) version 2.0 in simulated helicopter training.* Naval Postgraduate School.

Hoffman, R. R., & Palermo, D. S. (1991, 1991). *Cognition and the symbolic processes : applied and ecological perspectives*, Hillsdale, N.J.

Huemer, V. A., Hayashi, M., Renema, F., Elkins, S., McCandless, J. W., & McCann, R. S. (2005). Characterizing Scan Patterns in a Spacecraft Cockpit Simulator: Expert Versus Novice Performance. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 49*, 83-87.

Hyönä, J., Radach, R., & Deubel, H. (2003). *The mind's eye: cognitive and applied aspects of eye movement research*. Amsterdam; Boston: North-Holland.

Jungkunk, P. (2009). *Modeling Human Visual Perception for Target Detection in Military Simulations.* Naval Postgraduate School, Monterey.

Juno, K., Stephen, A. P., April, A., & Robert, S. A. (2010). Pilot gaze and glideslope control. *ACM Trans. Appl. Percept., 7*(3), 1-18.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist, 38*(1), 23-31.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351-371.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Wickens, C. (2001). *Comparision of expert and novice scan behaviors during VFR flight.* Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH.

King, G. W., & Lakey., C. B. (2006). *Cleared hot : a forward air control (airborne) concepts trainer* Naval Postgraduate School.

Kirwan, B., & Ainsworth, L. K. (1992). *A guide to task analysis*. London; Washington, DC: Taylor & Francis.

Klein, G. (2008). Naturalistic Decision Making. *Human factors : the journal of the Human Factors Society., 50*(3), 456.

Kulakowski, W. J. (2004). *Exploring the Feasibility of the Virtual Environment Helicopter System (VEHELO) for Use as an Instructional Tool for Military Helicopter Pilots.* Unpublished Unpublished Master's Thesis (I Think?), Naval Postgraduate School.

Law, B., Atkins, M. S., Kirkpatrick, A. E., & Lomax, A. J. (2004). *Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.

Lennerton, M. (2004). *Exploring a Chromakeyed Augmented Virtual Environment (ChrAVE) for Viability as an Embedded Training System for Military Helicopters.* Naval Postgraduate School.

Lesgold, A. M. (1983). *Acquiring expertise*. Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Lesgold, A. M., Feuer, M. J., & Black, A. M. (1997). Transitions in work and learning implications for assessment : papers and proceedings. from http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=965

Liu, Z.-Q., Yuan, X.-G., Fan, Y.-B., Liu, W., & Kang, W.-Y. (2009). Comparison of Expert and Novice Eye Movement Behaviors during Simulated Landing Flight *Space Medicine & Medical Engineering, 22*(5), 358-361.

Marshall, S. P. (2007). IV NEUROPHYSIOLOGIC MEASURES AND AUGMENTED COGNITION - Identifying Cognitive State from Eye Metrics. *Aviation space and environmental medicine., 78*(5), B165.

McLean, T. D. (1999). *An interactive virtual environment for training map-readings skill in helicopter pilots.* Naval Postgraduate School.

Merrill, M. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43-59.

Merrill, M. D. (1990). Second Generation Instructional Design (ID2). *Educational Technology, 30*(2), 7-14.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psycholigical Review, 63*, 81-97.

Ottati, W. L., Hickox, J. C., & Richter, J. (1999). Eye scan Patterns of Experienced and Novice Pilots During Visual Flight Rules (VFR) Navigation. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 43*, 66-70.

Phillips, J., McDermott, P. L., Thordsen, M., McCloskey, M., Klein, G., & Klein Associates Inc Fairborn, O. H. (1998). Cognitive Requirements for Small Unit Leaders in Military Operations in Urban Terrain. from http://handle.dtic.mil/100.2/ADA355505

Phillips, J. K., Shafer, J., Ross, K. G., Cox, D. A., & Shadrick, S. B. (2006). Behaviorally Anchored Rating Scales for the Assessment of Tactical Thinking Mental Models.

Platte, W. L., & Powers, J. J. (2007). *Using Motion Capture to Determine Marksmanship Shooting Profiles: Teaching Soldiers to Shoot Better Faster.* Naval Postgraduate School.

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual Span in Expert Chess Players: Evidence From Eye Movements. *Psychological Science, 12*(1), 48-55.

Ross, K. G., Phillips, J. K., Klein, G., & Cohn, J. (2005). *Creating Expertise: A Framework to Guide Simulation-Based Training* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2005, Orlando, Fla. U.S.A.

Russell, T. (1993). Reflection-in-Action and the Development of Professional Expertise. *Teacher Education Quarterly, 20*(1), 51-62.

Schmorrow, D., Cohn, J., & Nicholson, D. (2009). *The PSI handbook of virtual environments for training and education : developments for the military and beyond*. Westport, Conn.: Praeger Security International.

Schmorrow, D., Estabrooke, I., Grootjen, M., Campbell, G., Belz, C., & Luu, P. (2009). “What Was He Thinking?”: Using EEG Data to Facilitate the Interpretation of Performance Patterns. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (Vol. 5638, pp. 339-347): Springer Berlin / Heidelberg.

Schmorrow, D., Estabrooke, I., Grootjen, M., Coyne, J., Baldwin, C., Cole, A., et al. (2009). Applying Real Time Physiological Measures of Cognitive Load to Improve Training. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience* (Vol. 5638, pp. 469-478): Springer Berlin / Heidelberg.

Schmorrow, D., Reeves, L., Bolton, A., Campbell, G., & Schmorrow, D. (2007). Towards a Closed-Loop Training System: Using a Physiological-Based Diagnosis of the Trainee’s State to Drive Feedback Delivery Choices. In *Foundations of Augmented Cognition* (Vol. 4565, pp. 409-414): Springer Berlin / Heidelberg.

Schneider, W. (1985). Training High-Performance Skills: Fallacies and Guidelines. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 27*, 285-300.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (2000). *Cognitive task analysis*. Mahwah, N.J.: L. Erlbaum Associates.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Advances in child development and behavior, 10*, 9-55.

Sullivan, J. (1998). *Helicopter Terrain Navigation Training Using A Wide Field Of View Desktop Virtual Environment.* Naval Postgraduate School.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Tien, G., Atkins, M. S., Zheng, B., & Swindells, C. *Measuring situation awareness of surgeons in laparoscopic training*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.

Trowbridge, M. H., & Cason, H. (1932). An experimental study of Thorndike's theory of learning. *Journal of General Psychology, 7*, 245-258.

van Merriënboer, J., Clark, R., & de Croock, M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development, 50*(2), 39-61.

van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist, 38*(1), 5-13.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review, 17*(2), 147-177.

Waller, D., Hunt, E., & Knapp, D. (1998). The Transfer of Spatial Knowledge in Virtual Environment Training. *Presence: Teleoperators & Virtual Environments, 7*(2), 129-143.

Welch, G., & Davis, L. (2008). Tracking for Training in Virtual Environments: Estimating the Pose of People and Devices for Simulation and Assessment. In J. Cohn, D. Nicholson & D. Schmorrow (Eds.), *The PSI Handbook of Virtual Environments for Training and Education: Developments for the Military and Beyond*: Praeger Security International.

Whitehead, A. N. (1929). *The aims of education & other essays*. New York: Macmillan Co.

Wickens, C. D., Gordon, S. E., & Liu, Y. (2004). *An introduction to human factors engineering*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Wright, G. T. (2000). *Helicopter Urban Navigation Training using virtual environments* Naval Postgraduate School.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A

HIGH-LEVEL COGNITIVE TASK ANALYSIS OF ROTARY WING TACTICAL OVERLAND FLIGHT TO OBJECTIVE

This is the high-level representation only. Details of each component within this representation can be found in the following sections. Each of the primary sub-goals represented here: Complete-flight-planning-operations, Complete-pre-flight-operations, and Complete-in-flight-navigation-procedures is described in its own section to follow.

GOAL: Complete-rotary-wing-tactical-overland-flight-to-objective	; generic task description to include TRAP/CSAR/NEO/INFIL/EXFIL
GOAL: Complete-flight-planning-operations	; typically ready-room activities; navigation component of detailed mission planning, including time enroute, anticipated track and fuel required
GOAL: Acquire-navigation-materials	; often available digitally using JMPS
GOAL: Conduct-map-study	
GOAL: Conduct-map-preparation	; annotate maps with route and timing information
GOAL: Conduct-NVG-pre-operational-checks	; using NVG operator's manual
GOAL: Complete-pre-flight-operations	
GOAL: Configure-cockpit-for-navigation	; arrange maps and kneeboard checklist to facilitate rapid scan and effective navigation.
GOAL: Configure-aircraft-for-dual-ship- NVG-flight	
GOAL: Conduct-preflight-navigation-system-initialization	

GOAL: Complete-in-flight-navigation-procedures	
GOAL: Navigate-to-initial-point	
GOAL: Navigate-to-next-waypoint	; at each checkpoint, perform cockpit maintenance duties including a check of planned versus actual timelines.
GOAL: Maintain-orientation	; this is the basic default method, in absence of any higher priority task, PNAC attempts the best possible update of plotted position
GOAL: Adjust-speed-for-arrival-time	
GOAL: Adjust-course-if-required	
GOAL: Execute-Magellan-procedures	; procedures for lost aircraft, may involve mission abort
Repeat-until-complete	

PLANNING PHASE COGNITIVE TASK ANALYSIS

Overland <TRAP/CSAR/NEO/INFIL/EXFIL> Mission

Assume the pilot has been given specific mission objectives and constraints to include aircraft configuration, crew load, area of operation, and mission support.

First primary objective is to complete the planning phase of the task. This involves acquiring maps, aerial photos, intelligence data, etc. that will be used for planning flight paths, spider routes, and assumed accuracy and location of assumed threats.

GOAL: Acquire-navigation-materials	
SELECT: Map-study-method	
Acquire-correct-map	; multiple scales if available
Aerial-photo-method	
Acquire-aerial-photo	
Satellite-photo-method	
Acquire-satellite-photo	
Combined-method	; preferred method assuming all are available
Acquire-all-available-assets	; either paper or JMPS
GOAL: Conduct-map-study	
GOAL: Conduct-legend-study	; study the legend for all specifics to be used in next phase
Determine-horizontal-scale	
Determine-elevation-scale	
Determine-units	; in meters, feet. etc.
Calculate-conversion	; to bring into aircraft units
Determine-contour-interval	
Determine-vegetation-types	

Determine-cultural-features	
Determine-populous-areas	; high intensity lighting makes NVG use difficult in vicinity of populous areas
Determine-magnetic-variation	
GOAL: Conduct-detail-map-study	; pre-route planning activity
Locate-threats	; based on current intelligence (JMPS)
Plot-threats	
Locate-area-of-interest	; e.g. landing zone
Plot-area-of-interest	
Locate-current-flight-hazards	; e.g. power lines, (JMPS/ECHUM)
Plot-current-flight-hazards	
Determine-SAFE-areas	
Plot-SAFE-areas	
Compute-threat-areas	; JMPS
GOAL: Analyze-threat-envelopes	
Rotary-wing-threat-envelope	
Fixed-wing-threat-envelope	
GOAL: Analyze-terrain-features	; based on what is available
SELECT: Prominent-recognizable-checkpoints-method	
Prominent-limiting-features-method	
Prominent-guiding-features-method	
Combination-method	; always the preferred method
GOAL: Analyze-NVG-flight-considerations	; if the mission will/could be flown under NVG conditions
Checkpoint-analysis	; ensure key features can be identified under NVG/low lighting conditions
Avoid-flying-directly-toward-light-sources	; use doglegs to avoid flying directly at high intensity lighting (i.e., cities, moon)

Analyze-moon-position-and-angle	; consider effects of shadows
GOAL: Select-navigation-points-for-primary-ingress-route	; navigation fixes (turn points along route) are selected from cue list
Calculate-distance-of-ingress-route	
Calculate-time-of-ingress-route	
Calculate-fuel-for-ingress-route	
GOAL: Annotate-map-&-kneeboard-card	
Anticipated-track	
Anticipated-progress-interval-marks	;'tick' marks to be used in flight to judge progress along track; useful for estimating terrain features that should be in view at any particular time along route
User-specific-navigation-aids	;e.g. highlighting specific contour intervals
Doghouse-information	;for each leg, maps are normally annotated using a doghouse shaped box for each route leg. This information includes heading to next checkpoint, groundspeed, fuel ladder information and time to next checkpoint,
GOAL: Select-navigation-points-for-secondary-ingress-route	
Calculate-distance-of-ingress-route	
Calculate-time-of-ingress-route	
Calculate-fuel-for-ingress-route	
GOAL: Annotate-map-&-kneeboard-card	
Anticipated-track	
Anticipated-progress-interval-marks	
User-specific-navigation-aids	
Doghouse-information	
GOAL: Select-navigation-points-for-	; egress route will normally be different from

primary-egress-route	ingress route to minimize the likelihood the enemy forces alerted during ingress will have an opportunity to respond
Calculate-distance-of-egress-route	
Calculate-time-of-egress-route	
Calculate-fuel-for-egress-route	
GOAL: Annotate-map-&-kneeboard-card	
Anticipated-track	
Anticipated-progress-interval-marks	
User-specific-navigation-aids	
Doghouse-information	
GOAL: Select-navigation-points-for-secondary-egress-route	
Calculate-distance-of-egress-route	
Calculate-time-of-egress-route	
Calculate-fuel-for-egress-route	
GOAL: Annotate-map-&-kneeboard-card	
Anticipated-track	
Anticipated-progress-interval-marks	
User-specific-navigation-aids	
Doghouse-information	
Calculate-mission-timeline	;mission timeline generally uses the longest anticipated ingress and egress routes, any loiter time, and time to complete mission (e.g. land and pick up troops.)
Calculate-total-fuel-required-for-mission	
GOAL: Compare-required-fuel-with-maximum-gross-weight	;required fuel includes NATOPS, typewing, airwing and squadron mandated reserves
Adjust-mission-configuration	;if the mission requires more fuel than can be carried due to gross weight constraints,

	either the navigation route (and fuel requirements) must be reduced, or the aircraft configuration (ordnance and crew) must be adjusted.
GOAL: Account-for-fuel-in-route	;if the mission can not be completed with adequate fuel reserves the navigation route or mission requirements must be updated.
Adjust-navigation-route	
Adjust-mission-configuration	
GOAL: Prepare-in-flight-guides	
GOAL: Prepare-kneeboard-cards	; possibly generated by JMPS
Prepare-communication-cards	
Prepare-brevity-code-words	
Prepare-strip-charts	; possibly generated by JMPS
GOAL: Prepare-annotated-maps	
Load-data-points-in-tactical-navigation-computer-mission-data-loader	;PFPS has the capability of loading a set of waypoints directly into a Mission Data Loader

PRE-FLIGHT PREPARATION PHASE (AIRCRAFT CONFIGURATION TO TAKE-OFF) Overland <TRAP/CSAR/NEO/INFIL/EXFIL> Mission

Assume successful completion of planning phase tasks and all associated objectives. Second primary objective is to prepare the cockpit for the actual flight. This begins with the pre-flight preparation, and concludes with the aircraft in the air beginning the overland navigation component.

GOAL: Configure-cockpit-for-navigation	; required inflight reference material (maps and kneeboard cards) must be readily accessible
Configure-maps	; e.g. fold correctly
Configure-kneeboard-cards	
GOAL: Configure-aircraft-for-dual-ship-NVG-flight	
Check-external-lighting	
Attach-chem-lights-to-aircraft	; aircraft not configured with external NVG compatible lighting may use chem lights
GOAL: Conduct-preflight-navigation-system-initialization	
GOAL: Conduct-preflight-checks	
Check-navigation-computer	
Check-GPS	
Check-TACAN	; Tactical Air Navigation
Check-Doppler	; Inertial navigation system
Check-INS	
Check-RADALT	; Radar altimeter
Check-transponder	
Check-search-light	; for NVG
Check-cockpit-lighting	; for NVG
Check-compass-system	
Load-waypoints-in-navigation-computer	

Select-waypoints-to-create-primary-ingress-route	
Select-waypoints-to-create-secondary-ingress-route	
Enter-magnetic-variation-information-in-tactical-navigation-computer	
GOAL: Conduct-final-NVG-function-check	; verify adequate image quality and proper focus
GOAL: Conduct-post-takeoff-systems-checks	
Verify-navigation-equipment-operational	; from list above. Additionally, verify and align compass systems.
Conduct-navigation-to-initial-point	
Identify-ingress-point-on-map	
Identify-feature-to-aid-identifying-initial-point	; a nearby prominent landmark
Scan-field-of-view-for-navigation-aid	
Locate-navigation-aid	
Positively-identify-initial-point	
Estimate-arrival-time-at-initial-point	
Adjust-speed-to-arrive-at-ingress-point-on-time	
Adjust-course-to-overfly-initial-point	
Adjust-speed-to-arrive-at-ingress-point-according-to-timeline	
Use-visual-aids-to-identify-ingress-point	
Verify-ingress-point-with-cockpit-navigation-aids	
Select-waypoint-from-tactical-navigation-computer	
Execute-navigate-to-next-waypoint	

EXECUTION PHASE (IN-FLIGHT EXECUTION OF ROUTE)
Overland <TRAP/CSAR/NEO/INFIL/EXFIL> Mission

Assume successful completion of all preceding tasks and associated objectives. The last primary objective is the actual in-flight navigation component. Because we make no assumptions as to the length and duration of the flight, nor do we assume anything about the terrain in question, we assume a simple repeated procedure for each pre-planned leg of the flight. For each leg, the navigating pilot will conduct a number of sub-tasks involving orientation to the environment and self-location. Communication to the PAC (pilot-at-controls) is included. If disorientation occurs (or even if it is believed to have occurred), the sub-goal Execute-Magellan-procedure is entered which involves re-orienting and getting back on route.

GOAL: Navigate-to-next-waypoint	
Start-leg-timing	
GOAL: Direct-flying-pilot-to-predetermined-heading	;the method is selected based on time available and visual cues present. If there are fewer non-ambiguous landmark features in view, one of the more time consuming methods may be required. Additionally, if the PNAC cannot quickly identify and communicate a unique landmark feature, a more time consuming method may be required.
SELECT: Use-landmark-method	;e.g 'saddle to the right of the peak at your two o'clock'. This method has the advantage that it allows the flying pilot flexibility on *how* to get to the specified location. The flying pilot can proceed at his discretion with little further assistance; thus providing the pilot not at controls (PNAC) more time to devote to comparing terrain

	features to map representation.
Identify-discernable-feature	
Direct-PAC-to-feature	; PAC=Pilot at controls
Use-clock-position-method	;turns are relayed to the PAC using clock position calls (rather than heading) to minimize the inside scan requirements of the flying pilot.
Specify-heading-by-clock-position	;this method places higher demand on the PNAC than the landmark method. After the initial turn, the PNAC will need to update the PAC quickly. It gives the flying pilot little flexibility in controlling the route of flight.
Using-turn-&-rollout-calls-method	;this is the most demanding method for the non-flying pilot since in general it demands complete attention for the duration of the turn. Additionally, the information it provides to the PAC has the shortest duration. The PAC will require further guidance quickly.
Specify-series-of-specific-actions	
Adjust-navigation-needle-to-new-course	;the navigation needle is often used to provide both pilots a backup of the intended heading between fixes.
Check-timing	;at each checkpoint, the PAC should compare the anticipated time enroute with the actual time enroute
Record-deviation-in-timing	
GOAL: Adjust-timing	
SELECT: Late-arrival-method	i.e., time has passed and you're not there yet
SELECT: Low-confidence-in-navigation-solution-method	
Defer-adjusting-speed	; you might be lost
High-confidence-in-navigation-solution-	

method	
Estimate-initial-increment-in-speed-required	; fix timing and proceed
Direct-PAC-to-adjust-IAS	; IAS Indicated Airspeed
Early-arrival-method	; already over checkpoint before specified time
Estimate-initial-decrement-in-speed-required	
Direct-PAC-to-adjust-IAS	
Verify-PAC-proceeding-correctly	;after directing a change in speed, the PNAC needs to follow up to ensure the correct change has been applied
GOAL: Check-ground-speed	
Scan-cockpit-gauges	
SELECT: No-Correction-Method	
No-action-required	
Correction-method	
Direct-PAC-to-adjust-speed	
GOAL: Check-on-track-progress	; actual vs. planned
SELECT: Within-limits-method	
No-action-required	
Outside-limits-method	
Estimate-required-change-in-ground-speed-(delta-GS)	; to minimize inside scan requirements, the PNAC directs the PAC using indicated airspeed. The calculations for adjusting timing are based on ground speed.
GOAL: Calculate-new-IAS	
Scan-current-IAS	
Add-delta-GS-as-to-IAS	
Direct-PAC-to-new-adjusted-speed	

GOAL: Verify-heading-is-correct	;this is a two-part correction. The PNAC estimates (or, if available, scans cockpit instrumentation to acquire) the heading required to maintain track. The PNAC must first determine if the PAC is flying the intended heading, and then verify that the resultant track is correct.
Scan-gauges	
SELECT: On-heading-method	
No-action-required	
Off-heading-method	
GOAL: Correct-heading	
Direct-PAC-turn	
SELECT: Use-landmark-method	;refer to previous discussion concerning preferred method and resultant PNAC workload.
Identify-discernable-feature	
Direct-PAC-to-feature	
Use-clock-position-method	
Specify-heading-by-clock-position	
Using-turn-&-rollout-calls-method	
Specify-series-of-specific-actions	
GOAL: Verify-track-is-correct	
Scan-gauges	
SELECT: On-track-method	
No-action-required	
Off-track-method	
GOAL: Correct-heading	
Direct-PAC-turn	

SELECT: Use-landmark-method	
Identify-discernable-feature	
Direct-PAC-to-feature	
Use-clock-position-method	
Specify-heading-by-clock-position	
Using-turn-&-rollout-calls-method	
Specify-series-of-specific-actions	
GOAL: Determine-aircraft-position	
Scan-heading-&-track	
Align-map-with-aircraft-track	
Analyze-terrain-within-field-of-view	
SELECT: salient-navigation-cues-in-view-method	;see cue list for details on 'salient' cues
GOAL: Match-navigation-feature-with-map-representation	
Estimate-map-representation-of-salient-navigation-cues	
Compare-estimated-map-representation-of-features-in-view-with-map	
Locate-potential-match-on-map	
Compare-map-with-feature-to-verify	;this may involve scanning from world to map multiple times. If feature goes out of view, procedures starts over with determine-aircraft-position
SELECT: positive-match-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	

Ambiguous-match-method	
Analyze-terrain-for-correlating-feature	;a possibly ambiguous feature that because of it's spatial relationship with other features may be used to definitively locate aircraft position
SELECT: possible-correlating-feature-in-view-method	
Estimate-map-representation-of-correlating-feature	
Compare-estimated-representation-of-feature-with-map	
Compare-map-with-feature-to-verify	
SELECT: positive-match-of-correlating-feature-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	
No-positive-match-of-correlating-feature-method	
Fly-time-distance-heading	
Update-aircraft-position-on-map-based-on-time-distance-heading	
Continue-analyzing-and-comparing-until-found-or-lost	
No-possible-correlating-feature-in-view-method	
Fly-time-distance-heading	
Update-aircraft-position-on-map-	

based-on-time-distance-heading	
Continue-analyzing-and-comparing-until-found-or-lost	
Candidate-feature-positively-classified-as-misidentified	;based on further analysis, the feature selected from field of view is determined NOT to be the feature originally selected on the map
GOAL: determine-if-any-positive-match-can-be-made	;one of three cases will apply:
SELECT: feature-on-map-identified-elsewhere-in-field-of-view-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	
Feature-in-field-of-view-positively-identified-elsewhere-on-map-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	
No-match-found	
Fly-time-distance-heading	
Update-aircraft-position-on-map-based-on-time-distance-heading	
Continue-analyzing-and-comparing-until-found-or-lost	

GOAL: determine-if-lost
Update-expected-position-on-map-using-time-distance-heading
Align-map-with-aircraft-track
Analyze-map-for-prominent-feature-within-expected-field-of-view
Analyze-terrain-for-possible-match-with-prominent-feature
SELECT: match-found-method
Estimate-distance-&-bearing-to-feature
Estimate-position-on-map-based-on-distance-&-bearing-to-feature
Update-position-on-map
Plot-current-position
Determine-navigation-correction-required
SELECT: major-deviation-method
Determine-new-course-to-route
Treat-current-position-as-new-waypoint
Execute-navigate-to-waypoint
Minor-deviation-method
Execute-correct-track-error
No-match-found-method
Query-crew-for-salient-cues
SELECT: no-cue-provided-method
Execute-Magellan-procedure
Cue-provided-method
Query-crew-for-description-of-cue

SELECT: match-found-method
Query-crew-for-distance-&-bearing-to-feature
Estimate-position-on-map-based-on-distance-&-bearing-to-feature
Update-position-on-map
No-match-found-method
Execute-Magellan-procedure
Query-wingman-for-salient-cues
SELECT: no-cue-provided-method
Execute-Magellan-procedure
Cue-provided-method
Query-wingman-for-description-of-cue
SELECT: match-found-method
Query-wingman-for-distance-&-bearing-to-feature
Estimate-position-on-map-based-on-distance-&-bearing-to-feature
Update-position-on-map
No-match-found-method
Execute-Magellan-procedure
Maintain-orientation
GOAL: Scan-for-next-navigation-point ; see cue inventory
SELECT: Follow-hand-rail-method ; usually a linear terrain feature

Positively-identify-hand-rail-feature	
Direct-PAC-to-follow-hand-rail	
GOAL: Update-on-track-progress	
Select-on-track-landmark	
Evaluate-track-deviation	
Visible-intermediate-navigation-point-method	; only if you see the point
Direct-PAC-to-navigation-point	
Verify-PAC-proceeding-to-correct-feature	
SELECT: PAC-proceeding-to-correct-feature-method	
Continue-to-navigation-point	
PAC-not-proceeding-to-correct-feature-method	
Inform-PAC	
Direct-PAC-to-correct-navigation-point	
Continue-to-navigation-point	
Proceed-through-ambiguous-area-method	; no useful immediate cues, so fly on until this changes
Update-expected-position-on-map-using-time-distance-heading	
Align-map-with-aircraft-track	
Analyze-map-for-prominent-feature-within-expected-field-of-view	
Analyze-terrain-for-possible-match-with-prominent-feature	
Continue-until-match-found-or-lost	
Time-distance-heading-method	; always available, use dead-reckoning. PAC responsible for maintaining mean track, subject to error.
Update-expected-position-on-map-using-time-distance-heading	
Align-map-with-aircraft-track	
Analyze-map-for-prominent-feature-within-	

expected-field-of-view	
Analyze-terrain-for-possible-match-with-prominent-feature	
Continue-until-match-found-or-lost	
Execute-determine-accurate-current-location-steps	; essentially, ‘determine-accurate-current-location’ becomes a default action
Update-current-position-mark-on-map	; successful outcome of ‘determine-accurate-current-location’ is new position mark on map
Check-time-enroute	
Compare-time-en-route-with-progress-tick-mark-on-map	
Compare-position-on-map-with-plotted-track	
Estimate-horizontal-deviation	; locate yourself on the map and mark it
Estimate-impact-on-navigation-and-timing	
Estimate-impact-on-exposure	
SELECT: Course-correction-required-method	
GOAL: Correct-heading	
Direct-PAC-turn	
SELECT: Use-landmark-method	
Identify-discernable-feature	
Direct-PAC-to-feature	
Use-clock-position-method	
Specify-heading-by-clock-position	
Using-turn-&-rollout-calls-method	
Specify-series-of-specific-actions	
Speed-correction-required-method	
Direct-PAC-to-adjust-speed	
GOAL: Execute-Magellan-procedures	;the option to select will depend on following factors:
Analyze-current-terrain-for-threat	;how close are enemy forces presumed to be and what is the level of confidence in

	troop location information
Analyze-current-terrain-for-exposure	;if it is possible to climb without increasing exposure, increased altitude will afford more opportunity to find recognizable landmark
Analyze-current-terrain-for-signature	;hovering may not be an option if based on power required and fuel constraints signature will beacon enemy.
Analyze-timing-for-ahead	
Analyze-timing-for-behind	
Analyze-fuel-on-board-compared-to-estimated-fuel-required	;if fuel is near limits, landing will use less fuel. Since this will likely reduce terrain features in view, landing is only practical if help is available
Analyze-degree-of-confidence	
Analyze-potential-assistance-with-navigation	;is RESCORT/RESCAP available to help
SELECT: Confess-method	
SELECT: Wingman-method	; if wingman is available, is he disoriented also?
Initiate-radio-call	
RESCORT/RESCAP-method	
Initiate-radio-call	
Orbit-method	
Provide-orienting-feature-for-PAC	
Attempt-to-match-terrain-feature-with-map-representation	
SELECT: feature-recognized-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	
Plot-current-position	
Determine-navigation-correction-required	
SELECT: major-deviation-method	

Determine-new-course-to-route
Treat-current-position-as-new-waypoint
Execute-navigate-to-waypoint
Minor-deviation-method
Execute-correct-track-error
No-feature-recognized-method
Attempt-to-recognize-prominent-feature
Continue-until-abort-criteria-met-or-match-found
NOE-method
Provide-area-for-NOE-to-PAC
Attempt-to-match-terrain-feature-with-map-representation
SELECT: feature-recognized-method
Estimate-distance-&-bearing-to-feature
Estimate-position-on-map-based-on-distance-&-bearing-to-feature
Update-position-on-map
Plot-current-position
Determine-navigation-correction-required
SELECT: major-deviation-method
Determine-new-course-to-route
Treat-current-position-as-new-waypoint
Execute-navigate-to-waypoint
Minor-deviation-method
Execute-correct-track-error
No-feature-recognized-method
Attempt-to-recognize-prominent-feature
Continue-until-abort-criteria-met-or-match-found
Hover-method
Provide-direction-to-hover-area-to-PAC
Attempt-to-match-terrain-feature-with-map-

representation	
SELECT: feature-recognized-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	
Plot-current-position	
Determine-navigation-correction-required	
SELECT: major-deviation-method	
Determine-new-course-to-route	
Treat-current-position-as-new-waypoint	
Execute-navigate-to-waypoint	
Minor-deviation-method	
Execute-correct-track-error	
No-feature-recognized-method	
Attempt-to-recognize-prominent-feature	
Continue-until-abort-criteria-met-or-match-found	
Land-method	
Determine-appropriate-LZ	;see cue chart for LZ evaluation criteria
Direct-PAC-to-LZ	
Attempt-to-match-terrain-feature-with-map-representation	
SELECT: feature-recognized-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	
Plot-current-position	
Determine-navigation-correction-required	
SELECT: major-deviation-method	
Determine-new-course-to-route	
Treat-current-position-as-new-waypoint	

Execute-navigate-to-waypoint	
Minor-deviation-method	
Execute-correct-track-error	
No-feature-recognized-method	
Attempt-to-recognize-prominent-feature	
Continue-until-abort-criteria-met-or-match-found	
Climb-method	; nap of the Earth
Direct-PAC-to-new-altitude	
Attempt-to-match-terrain-feature-with-map-representation	
SELECT: feature-recognized-method	
Estimate-distance-&-bearing-to-feature	
Estimate-position-on-map-based-on-distance-&-bearing-to-feature	
Update-position-on-map	
Plot-current-position	
Determine-navigation-correction-required	
SELECT: major-deviation-method	
Determine-new-course-to-route	
Treat-current-position-as-new-waypoint	
Execute-navigate-to-waypoint	
Minor-deviation-method	
Execute-correct-track-error	
No-feature-recognized-method	
Attempt-to-recognize-prominent-feature	
Continue-until-abort-criteria-met-or-match-found	

CUE INVENTORIES

Environmental Cues

Identify unique features to correlate expected position with actual position -- scan outside, query crew. These are used in a repeated fashion throughout the flight, but are particularly triggered in re-orienting and map-correlation sub-tasks.

CUE	DESCRIPTION
Unique, distinguishable terrain feature based on three-dimensional shape and orientation. (Three dimensional shape infers use of altitude as correlating feature.)	A key characteristic of a terrain feature to be used as a navigation checkpoint is that it is uniquely identifiable. Navigation routes are planned such that, when practical, such a terrain feature is always in view.
Unique, distinguishable cultural feature	Cultural features are considered secondary navigation aids. Flying in close proximity to cultural features generally increases the exposure to enemy forces. Distant cultural features visible from long ranges and low altitudes (i.e., poles for power lines, water towers) are more commonly used than terrain features that would be associated with dense population areas (towns, highways and rivers.) The accuracy of depicted cultural features often relates to the likelihood of exposure to enemy forces. (Compare jeep trails with hardball roads.)
Distinguishable location based on relation and orientation of two or more non-distinct terrain features	If a single unique terrain feature cannot be selected, position may be determined by using the spatial relationship (distance and orientation) of more than one non-distinct terrain feature. This is considered a lower priority since it relies on keeping multiple features within the field of view (or coordinating with crewmembers.) Given cockpit visibility constraints, usually the time when multiple features are in view will be considerably less than the time a single feature is in view.

CUE	DESCRIPTION
Any discernible difference of terrain along selected navigation route from surrounding terrain	In areas with little terrain relief, navigation may rely on subtle variations in the terrain selected for the navigation route. For example in desert terrain, it may be appropriate to navigate along a dry creek bed or wash.
Any charted and discernible difference of vegetation along selected navigation route.	Lacking other navigational cues, vegetation can sometimes be used as a cue. For example, in areas with little terrain relief but ample coverage by vegetation, waterways will often be visible based on the difference in vegetation along the waterway.

Cockpit Cues

These are cues specific to inside the cockpit to include gauges and controls.

CUE	DESCRIPTION
Altitude above mean sea level (MSL) – barometric altitude	Elevation of terrain features is used as an identifying characteristic. Current altitude must be known to do this. Additionally, pilots will need to judge height of terrain relative to aircraft. (I.e. peak at aircraft ten o'clock is 200 feet above aircraft. Aircraft is at 1400' MSL. Peak is approximately 1600' MSL.)
Altitude above ground level altitude (AGL) – radar altitude.	Pilot will use current altitude to judge distance to objects.
Heading	Magnetic heading, also depending on the aircraft, true heading information may be available. Aircraft may also have a selectable navigation marker ('bug') that can be dialed to heading to fly. If the aircraft is equipped with a navigation computer a needle pointing to the next selected waypoint may also be available.
Track	If available in the aircraft, a track needle should be available to verify aircraft is on correct heading to maintain planned track.
Clock	Used to track total time enroute as well as individual navigation leg timing. Essential for time/distance/heading mode.
Ground speed	Required for PNAC to calculate maintenance of and correction to timeline.
Indicated airspeed (IAS)	Primary scan for PAC. Required for PAC to maintain airspeed to aid time/distance/heading calculations.

CUE	DESCRIPTION
Attitude indicator	Improve situational awareness and facilitates rapid scanning for PNAC. For example, PNAC can initiate a turn and then track progress of turn while checking cockpit gauges or map.
Current fuel onboard	If too much fuel is used on route or expected delay times are exceeded, the navigation route may need to be changed. Additionally, the procedures to follow if pilot is lost depend on fuel.
Turn rate	Useful for judging time required to complete a turn.

Landing Zone Cues

These are cues specific to a landing zone. These should be considered in addition to the Environmental Cues listed earlier.

CUE	DESCRIPTION
Size	Pilot must be able to determine if aircraft will be able to safely land and takeoff
Slope	Pilot must be able to determine (or approximate) if the slope of the terrain is within aircraft landing limits.
Suitability	Factors such as muddy or badly rutted landing areas and foreign object damage (FOD) hazards may make landing impractical.
Wind	Pilot must be able to determine wind direction. This can be done with cockpit instrumentation (comparing airspeed and groundspeed) or visual aids (direction of dust and smoke, movement of vegetation). Pilot must also be able to judge the effect of surrounding obstacles on wind (turbulence and loss of effect.)
Escape routes	Pilots must judge if an approach and departure path based on current winds can be safely executed with an acceptable margin of error and preserving a waveoff capability.
Elevation	Pressure and density altitude are required to determine if adequate power margin exists to safely conduct and approach, landing and takeoff after troop embarkation.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B

Training Expert Navigation Skills: Navigation Exercise Evaluation
Experiment Checklist

Date:

Subject ID:

- ☐ Email confirmation sent:
  - Subject show time:
  - Email sent (date/time):
- ☐ Notify lab participants of data collection time
- ☐ Validate equipment hardware and software
  - Screen brightness and contrast settings
  - Lab lighting conditions
- ☐ "Experiment in Progress" signs
- ☐ Introductory Script
- ☐ Informed Consent
- ☐ Visual Acuity equipment
- ☐ Spatial Aptitude test
- ☐ Background questionnaire
- ☐ Map set up
- ☐ Route brief
- ☐ Trial period instructions
- ☐ Calibration script
- ☐ Video recording equipment (storage media, files naming and backup scheme)
- ☐ Navigation exercise
- ☐ Save and backup data; folder name: subject ID and date
- ☐ Post-navigation exercise questionnaire
- ☐ Wrap up and thank you, contact information

Training Expert Navigation Skills: Navigation Exercise Evaluation

Experiment Checklist
6 August 2010

Date:

Subject ID:

Training Expert Navigation Skills: Navigation Exercise Evaluation
Welcome Script

Scheduled Arrival Time:

Actual Arrival Time:

End Time:

Hello and welcome. Thank you for participating. We hope that your participation will ultimately lead to improvements in our understanding of how pilots train for overland navigation. This research may also help us understand how we build and evaluate training simulations in general.

Today we'll be asking you to complete a short navigation exercise using a pc-based simulation. Before and after the navigation exercise we'll ask you to fill out some short questionnaires related to your background and experience. We'll have some brief vision and spatial abilities tests. During the navigation exercise we'll be using a system of cameras and software that record your eye movement.

We hope to take less than an hour. We ask for uninterrupted participation. During the simulation exercise and when near equipment please observe no food/drink restrictions. If you need to use a restroom they are located across the breezeway, through the double doors and to the left. Bottled water is available in the fridge by the door.

Are you ready to go on?

The next step is to make sure you understand any risks, the voluntary nature of participation and our efforts to protect your privacy.

Date:

Subject ID:

Naval Postgraduate School
Consent to Participate in Research

Introduction.

You are invited to participate in a research study entitled "Training Expert Navigation Skills". This project is focused on simulation tools for understanding helicopter overland navigation skills.

Procedures.

We are asking helicopter pilots to conduct a brief navigation exercise in a simulated environment. During the study we will use cameras and software to estimate and record eye gaze direction. The experiment consists of a short questionnaire and a multiple choice test of spatial ability. We will give you a map with a short route depicted and provide time for map study. We will then take a few minutes to adjust the cameras and software. After a familiarization period with the simulated environment you will have the opportunity to fly the route. During this time we will collect eye-scan information for later analysis. Following the navigation part of the study we will ask a few short questions about your experience. We will provide information on the study results. The entire test should take no more than 60 minutes. Any information will be used only for this research endeavor.

Voluntary Nature of the Study.

Your participation in this study is strictly voluntary. If you choose to participate you can change your mind at any time and withdraw from the study. You will not be penalized in any way or lose any benefits to which you would otherwise be entitled if you choose not to participate in this study or to withdraw.

Potential Risks and Discomforts.

The potential risks of participating in this study are minimal. You will be sitting in front of a large monitor using a conventional joystick to fly through the simulation environment. The eye scan camera systems have infrared lights associated with them. These are no more harmful than normal room lighting.

Anticipated Benefits.

Anticipated benefits from this study are an improved understanding of the task of helicopter overland navigation and the implications for design of training simulation tools.

Compensation for Participation.

No tangible compensation will be given. A copy of the research results will be available at the conclusion of the experiment. We will write up a short summary for later inclusion in larger research reports. All reports will be available on line or on request.

Confidentiality & Privacy Act.

Any information that is obtained during this study will be kept confidential to the full extent permitted by law. All efforts, within reason, will be made to keep your personal information in your research record confidential but total confidentiality cannot be guaranteed. All references to data collected will be made

anonymous. Your name will be encoded as a participant number. Only principle investigators will have access to this key that translate an identification number to your name. However, it is possible that the researcher may be required to divulge information obtained in the course of this research to the subject's chain of command or other legal body.

If you consent to be identified by name in this study, any reference to or quote by you will be published in the final research finding only after your review and approval. If you do not agree, then you will be identified broadly by discipline and/or rank, (for example, "fire chief").

☐ I consent to be identified by name in this research study.

☐ I do not consent to be identified by name in this research study.

Points of Contact.

If you have any questions or comments about the research, or you experience an injury or have questions about any discomforts that you experience while taking part in this study please contact the Principal Investigator, CDR Joseph Sullivan, 831-656-7582, sullivan@nps.edu. Questions about your rights as a research subject or any other concerns may be addressed to the Navy Postgraduate School IRB Chair, Dr. Angela O'Dea, 831-656-3966, alodea@nps.edu.

Statement of Consent.

I have read the information provided above. I have been given the opportunity to ask questions and all the questions have been answered to my satisfaction. I have been provided a copy of this form for my records and I agree to participate in this study. I understand that by agreeing to participate in this research and signing this form, I do not waive any of my legal rights.

____ ____________________
Participant's Signature Date

____ ____________________
Researcher's Signature Date

Date:

Subject ID:

Training Expert Navigation Skills: Navigation Exercise Evaluation
Background Questionnaire

1. Please provide the following information: Age:

   Gender:

2. To what extent have you participated in other activities other than helicopter overland navigation that may contribute to improved navigation skills? (Examples may include sport orienteering, land navigation exercises, participation in boy/girl scouts etc.)?

O	O	O	O	O
No Related Experience	Very Limited Related Experience	Limited Related Experience	Somewhat Significant Experience	Significant Related Experience

Please Describe:

3. At your peak of currency, how would you rate your navigation skills in a low-level (below 200' AGL) overland environment?

O	O	O	O	O
Poor	Fair	Average	Good	Excellent

4. If tasked today, how would you rate your navigation skills in a low-level (below 200' AGL) overland environment?

O	O	O	O	O
Poor	Fair	Average	Good	Excellent

5. How much experience do you have with low level navigation in mountainous desert terrain?

O	O	O	O	O
None	Very Little	Moderate	Considerable	Extensive

6. How much low level navigation experience do you have in the 29 Palms area?

O	O	O	O	O
None	Very Little	Moderate	Considerable	Extensive

Training Expert Navigation Skills: Navigation Exercise Evaluation

Background Questionnaire
6 August 2010

Training Expert Navigation Skills: Navigation Exercise Evaluation
Background Questionnaire

Flight Time Summary

Total Hours:	
Overland Hours:	
How long has it been since your last flight?	
How long has it been since your last overland navigation flight?	
Branch of Service:	
Community:	
Type aircraft flown:	
Years of aviation experience:	
Describe your operational flying experience:	

Date:

Subject ID:

Training Expert Navigation Skills: Navigation Exercise Evaluation
Equipment Familiarization Script

- Verify equipment on and running. Set up for 'Trial' mode.
  - Facelabs systems
  - IG PC
  - Display contrast and brightness settings
  - Instructor/data collection PC
  - Video recording equipment
  - Eye height calibration
- Check lab lighting and 'experiment in progress' signs posted.

Please have a seat. (Experimenter's seat.)

Before we start the navigation exercise, we'll let you get familiar with the system and controls. I'll briefly explain the set up and let you fly a sample navigation route to get familiar with the displays and controls.

The system provides an out the window view on the screen in front of you and an instrument cluster and map display on the monitor to your right. The joystick controls the aircraft position and the view. The aircraft will fly at a constant 60 knots ground speed. Altitude will be fixed near a constant 150' AGL. The lateral (roll) axis of the joystick is used to control heading. Pushing left or right will enter a coordinated turn in the appropriate direction. Radius of turn is roughly what you would expect in the aircraft. Releasing the joystick the aircraft will return to wings level. The longitudinal (pitch) axis controls the pitch angle of the view. To look down push forward; to look up pull back.

The map rotates automatically to align with the aircraft's heading. The map will also scroll automatically. The instrument cluster contains a compass that displays True Heading, a Barometric Altimeter a Radar Altimeter, and a digital clock that indicates elapsed time.

Before we give the controls a try, do you have any questions?

<...>

Please have a seat in the operator's station.

<...>

Now let's fly a practice route. The practice route is depicted on the map and is similar to route you will be flying. The aircraft is You'll now have a few minutes to get familiar with the controls. When you're comfortable, please fly to the waypoint. We'll pause there and then practice flying the route while exercising map and clock controls.

<...>

Training Expert Navigation Skills: Navigation Exercise Evaluation
Equipment Familiarization Script

You'll now have a few minutes to fly along the depicted route to get comfortable with the controls and display. This route consists of 3 legs, each about a minute and a half long. When the route is complete you can repeat it if you would like more practice. Let me know when you are ready and we'll start the practice route along the depicted path.

Thank you. Do you have any questions before we move on?

<...>

During the next stage we'll calibrate the eye-tracking equipment. Following this we'll provide a brief on the route you will be flying, provide some time for map study and then let you fly the test route.

Date:

Subject ID:

Training Expert Navigation Skills: Navigation Exercise Evaluation
Eye Scan Calibration Script

- Verify equipment on and running. Set up for 'Trial' mode.
    - Facelabs systems
    - IG PC
    - Display contrast and brightness settings
    - Instructor/data collection PC
    - Video recording equipment
    - Eye height calibration
- Check lab lighting and 'experiment in progress' signs posted.

We'll now calibrate eye tracking equipment. This should only take a few minutes.

Look straight ahead at the monitor in front of you with a neutral expression.

<take picture, create head model>

Please look directly into camera one.

Please look directly into camera two.

Next we need to calibrate the system. During this portion, you'll see a series of dots presented on the screen. Try to focus directly on the spot without blinking. Please let me know when you are ready and we'll continue…

<OTW screen calibration, save with subject ID number>

Look straight at the instrument panel/map display with a neutral expression.

<take picture, create head model>

Please look directly into camera three.

Please look directly into camera four.

Next we'll calibrate the screen used for the map and instrument display. As before, please focus on the dot without blinking. Please let me know when you are ready to continue…

<Instrument Panel calibration>

Date:

Subject ID:

Training Expert Navigation Skills
Navigation Route Difficulty Estimation

1. For each leg on the route, please estimate how difficult you think it will be to navigate by referencing terrain features. Place an 'X' on the line at the position that best indicates your estimate.

Leg One

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Two

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Three

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Four

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Five

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Six

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Training Expert Navigation Skills
Navigation Route Difficulty Estimation

Leg Seven

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Eight

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Nine

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Ten

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Eleven

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Training Expert Navigation Skills

Navigation Route Difficulty Estimation
6 August 2010

Date:

Subject ID:

Training Expert Navigation Skills
Navigation Route Difficulty Assessment

1. For each leg on the route, please estimate how difficult it was to navigate by referencing terrain features. Place an 'X' on the line at the position that best indicates your estimate.

Leg One

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Two

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Three

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Four

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Five

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Six

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Training Expert Navigation Skills
Navigation Route Difficulty Assessment

Leg Seven

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Eight

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Nine

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Ten

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

Leg Eleven

Completely Trivial	Somewhat Difficult	Moderately Difficult	Very Difficult	Not At All Possible

Comments:

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Mr. John Moore
   Navy Modeling and Simulation Office
   Washington, D.C.

4. Dr. Rudy Darken
   Naval Postgraduate School
   Monterey, California

5. LCDR Joseph Cohn
   Office of Naval Research
   Alexandria, Virginia

6. Dr. Anthony Ciavarelli
   Naval Postgraduate School
   Monterey, California

7. Dr. Theodore Lewis
   Naval Postgraduate School
   Monterey, California

8. Dr. Michael McCauley
   Naval Postgraduate School
   Monterey, California

9. Dr. Gurminder Singh
   Naval Postgraduate School
   Monterey, California